

Janne Antikainen – Jarmo Eskelinen – Heli Koski – Tommi Niemi –
Mika Pajarinen – Sinikukka Pyykkönen – Marc de Vries

Massadatasta liiketoimintaa ja tehokkaita julkisia palveluja

Huhtikuu 2016

Valtioneuvoston selvitys-
ja tutkimustoiminnan
julkaisusarja 16/2016

ISSN PDF 2342-6799

ISBN PDF 978-952-287-250-0

Tämä julkaisu on toteutettu osana valtioneuvoston vuoden 2015 selvitys- ja tutkimussuunnitelman toimeenpanoa (www.vn.fi/TEAS).

Julkaisun sisällöstä vastaavat tiedon tuottajat, eikä tekstisisältö edusta valtioneuvoston näkemystä.



Esipuhe

Olemme toteuttaneet raportoidun selvityksen Valtioneuvoston kanslian valtioneuvoston päätöksentekoa tukevan selvitys- ja tutkimustoiminnan määrärahoilla rahoitetun ”Massadatasta liiketoimintaa ja tehokkaita julkisia palveluita” -hankkeen puitteissa. Heli Koski (ETLA) on kirjoittanut Mika Pajarisen (ETLA) kanssa luvun 2 yrityskyselyyn perustuvan katsauksen massadatan käytöstä suomalaisissa yrityksissä. Jarmo Eskelinen valottaa luvussa 3 massadataekosysteemin kehittämistä Suomessa. Janne Antikainen (MDI), Sinikukka Pyykkönen (MDI) ja Jarmo Eskelinen (Forum Virium) ovat kirjoittaneet luvun 4 koskien massadatan hyödyntämistä julkisissa palveluissa. Luvussa 5 Marc de Vries (Spark Legal Network) ja Tommi Niemi (Business Law Finland) käsittelevät massadatan hyödyntämisen oikeudellisia näkökulmia tapaustutkimusten valossa.

Selvityksen tekemistä on ohjannut liikenne- ja viestintäministeriön, opetus- ja kulttuuriministeriön, työ- ja elinkeinoministeriön, valtiovarainministeriön ja ympäristöministeriön edustajista koostuva ohjausryhmä, jota kiitämme hyvistä keskusteluista ja arvokkaista kommentteista. Selvityksessä esitetyt tulokset, näkemykset ja johtopäätökset ovat kuitenkin tutkimusryhmän omia, eivätkä välttämättä heijasta työn ohjaamiseen osallistuneiden tahojen näkemystä.

Tekijät

Sisällys

Esipuhe	3
Sisällys	5
Tiivistelmä	7
1 Johdanto	9
2 Massadatan käyttö ja liiketoimintapotentiaali suomalaisissa yrityksissä	11
2.1 Suuret yritykset massadatan käytön eturintamassa	11
2.2 Massadatan liiketoimintapotentiaali: parempia päätöksiä, tuotteita ja asiakaskokemuksia	15
2.3 Massadatan hyödyntämisen rajoitteet: tarpeita ei nähdä, tietotaitoa puuttuu	17
2.4 Miten massadatan hyödyntäjät poikkeavat muista yrityksistä?	18
2.5 Päätelmät ja suositukset	19
3 Massadata: nykykäytännöt ja tulevat kehityssuunnat	20
3.1 Talouden dataistuminen: Massadataan liittyvät käsitteet ja riskit	20
3.2 Massadatan tulevat kehityssuunnat	22
3.3 Massadatastrategia	23
3.4 Omadata- ja blockchain-teknologiat	28
3.5 Päätelmät massadatan hyödyntämisestä	29
4 Massadatan hyödyntäminen julkisissa palveluissa	31
4.1 Massadatan lähteet ja tavoitteet massadatan käytölle julkishallinnossa	31
4.2 Kaupungit ja massadatan hyödyntäminen	32
4.3 Liikenne ja liikkuminen massadatan hyödyntäjänä	34
4.4 Hyvinvointipalvelut massadatan hyödyntäjinä	35
4.5 Avoimesta datasta datastrategiaan	36
4.6 Kansainvälinen yhteistyö	40
5 Innovatiivisia ratkaisuja massadatan hyödyntämiseen	41
5.1 Esimerkkien valinta ja lyhyt esittely	41
5.2 Henkilötietoja sisältävän datan uudelleenkäytön työkalu	41
5.3 Henkilötietoja sisältävän datan jakaminen luotetussa ympäristössä	44
5.4 Avoin julkinen varainkäyttö ja tietosuoja (Supervizor)	47
5.5 Päätelmät laajemmassa kontekstissa	49
5.6 Suositukset: esimerkkityökalujen sovellettavuus Suomeen	50

6 Johtopäätökset ja suositukset	51
6.1 Johtopäätökset	51
6.2 Suositukset	51
Lähteet	53
Liitteet	54
Liite 1 Yrityskysely: kuvia	54
Liite 2 Lista esimerkeistä koskien innovatiivisia ratkaisuja massadatan hyödyntämisessä	60



Tiivistelmä

Tässä raportissa selvitetään massadatan käyttöä ja arvioidaan sen liiketoimintapotentiaalia eri toimialoilla sekä kartoitetaan massa- ja omadatan sovelluspotentiaalia julkisen sektorin palveluissa sekä näihin palveluihin liittyviä oikeudellisia reunaehjoja. Tutkimushankkeen tarkoituksena oli tuottaa tietoa, jota voidaan käyttää digitaalisen kasvu ympäristön rakentamiseen tähtäävän hallituksen kärkihankkeen toimissa. Kärkihankkeen tavoitteena on lisätä massadatan hyödyntämistä ja käynnistää omadata-kokeiluja. Lisäksi tutkimushanke tukee EU:n henkilötietoasetuksen pohjalta uudistumassa olevan tietosuojalainsäädännön kansallista valmistelua tunnistamalla tiedon käyttöön liittyviä oikeudellisia reunaehjoja.

Raportti valottaa ensinnäkin kyselytutkimuksella kerätyn aineiston avulla sitä, kuinka paljon ja millä tavoin suomalaisissa yrityksissä ja eri toimialoilla hyödynnetään massadataa ja mikä on massadatan hyödyntämisen liiketoimintapotentiaali. Toiseksi, raportissa esitetään kirjallisuuskatsauksen, haastatteluiden ja työpajojen perusteella yksityisten ja julkisten massadataa hyödyntävien palveluiden kehittämiseen liittyviä ongelmakohtia ja pullonkauloja Suomessa. Raportissa tarkastellaan keinoja niiden ratkaisemiseksi ja massadataekosysteemin kehittämiseksi. Kolmanneksi, raportti esittelee kansainvälisiä, potentiaalisesti Suomeen soveltuvia innovatiivisia ratkaisuja, joilla massatietoihin perustuvien palveluiden kehittämistä voidaan edistää niin, että samalla turvataan riittävä tietosuoja.

Parempia päätöksiä, tuotteita ja asiakaskokemuksia

Marraskuussa 2015 toteutettuun suomalaisyritysten massadatan käyttöä kartoittavaan yrityskyselyyn vastasi 1 189 yritystä. Lähes neljäsosa kyselyyn vastanneista yrityksistä keräsi massadataa tai käytti sitä liiketoiminnassaan. Energia-alalla toimivat yritykset sekä rahoitus- ja vakuutustoimintaa harjoittavat yritykset ovat massadatan hyödyntämisen etujoukoissa. Kyselyn mukaan keskisuuret ja isot yritykset olivat selkeästi todennäköisemmin massadatan hyödyntäjiä kuin pienemmät yritykset, ja näiden joukossa odotettiin myös voimakkainta massadatan hyödyntämisen liikevaihto-osuuden kasvua. Lähes kolmannes massadataa käyttävistä yrityksistä arvioi massadatan hyödyntämisen vaikuttavan paljon tai hyvin paljon niiden liikevaihdon kasvuun lähitulevaisuudessa.

Tärkeimpiä massadatan hyödynnettyjä tietoja olivat yrityksen myyntijärjestelmistä ja yrityksen www-sivujen kautta kerätty data. Massadataa hyödynnettiin eniten päätöksenteossa, asiakas- ja markkina-analyysissa ja uusien tuotteiden kehittämisessä. Massadatan hyödyntämisen liiketoimintapotentiaalin arvioitiin realisoituvan yleisimmin parempien, tietoon perustuvien päätösten kautta, yrityksen kilpailuaseman ja tuotteiden laadun sekä asiakaskokemuksen paranemisen kautta.

Massadatan hyödyntämisen eturintaman yritykset odottavat merkittävää massadatan strategiseen hyödyntämiseen perustuvaa kasvua, mutta isossa joukossa suomalaisia yrityksiä ei massadatan käytön liiketoimintapotentiaalia kuitenkaan vielä tällä hetkellä ymmärretä. Myös massadatan käsittelyyn ja analyysiin liittyvä tietotaidon puute rajoittaa tai estää massadatan hyödyntämistä merkittävästi. Ensiarvoisen tärkeitä toimenpiteitä massadatan liiketoimintapotentiaalin valjastamiseksi ovat riittävien resurssien suuntaaminen datan käsittelyyn, hallintaan ja analytiikan sekä datatalouden yhteiskunnallistaloudelliseen koulutukseen ja massadatan hyödyntämistä edistävään tutkimus- ja innovaatiotoimintaan.

Omadata keskiössä massadataekosysteemin kehittämisessä

Yritysten ja julkisten tahojen haastatteluista ja heille suunnatuista työpajoista saatiin samanlainen viesti kuin yrityskyselyistä: Massadatan suurimmaksi hyödyksi ja tulevaisuuden mahdollisuudeksi nähdään potentiaali lisätä ja jalostaa tietoa päätöksenteon tueksi. Massadatan hyödyntämisessä tärkeänä pidetään tiedon käytettävyyttä, datan keruupaikkojen ja -välineiden laajentumista ja reaaliaikaistumista, datan määrän kasvun kautta syntyvää ennustuspotentiaalia sekä datalajien yhdistämisen ja kokemusdatan hyödyntämistä. Datan käytön haasteina nähdään mm. datamassan suuruus, maksullisuuteen tai maksuttomuuteen liittyvät kysymykset, omistajuus- ja pääsykysymykset, tiedon väärinkäytön ja katoamisen riski sekä tiedon antajien oikeuksien epäselvyydet.

Toimijoiden arvioissa massadatan liiketoimintapotentiaalista on eroavaisuuksia. Yritykset näkevät liiketoimintapotentiaalia datan keruussa ja analysoinnissa sekä tulosten hyödyntämisessä. Tutkimus- ja kehitystahot (esim. yliopistot ja tutkimuslaitokset) tarkastelevat massadatan käytön kysymyksiä sekä käytännöllisestä että teoreettisesta näkökulmasta. Kaupungit näkevät pääasialliseksi tehtäväkseen massadatan käytön mahdollistamisen ja datan keräämisen, muokkaamisen ja jakamisen avointen rajapintojen kautta. Tällä hetkellä massa- ja omadatan potentiaalia nähdään mm. liikenteessä ja omien terveyspalveluiden seurannassa. Liikenteen alalle on luotu viime vuosina paljon erilaisia reaaliaikaiseen ja staattiseen dataan perustuvia palveluita. Käyttäjän näkökulmasta terveyspalveluissa nousussa ovat omaan terveyteen liittyvät, monia datalajeja yhdistelevät palvelut.

Omadata on kriittisessä asemassa massadataekosysteemin kehittämisessä. Yksilöistä ja yksilöiden toiminnasta koottua dataa voi sisältyä lähes mihin tahansa massadataan. Sen huono hallinta voi hidastaa verkkopohjaisten palvelujen kasvua ja massadatasta toivottujen hyötyjen saamista. Haasteiden ratkaisemiseen tarvitaan uusia toimintamalleja, missä painotus siirtyy lainsäädännöstä käyttäjien hallitsemiin datan hallinnan työkaluihin. Yksi lupaavimmista mahdollisuuksista omadatatyökalujen kehittämiseen ovat blockchain/lohkokeiju -teknologiat, jotka sopivat hajautettuun tiedon varmentamiseen ja käsittelyyn, turvaratkaisuihin sekä arvon tai resurssien jakamiseen.

Innovatiivisia työkaluja julkiselle sektorille massadatan hyödyntämiseen

Mahdollisuus luoda, varastoida, analysoida ja jakaa valtavia määriä dataa silmäläpäkseen minimaalisin kustannuksin nostaa esille tietosuojaan liittyviä kysymyksiä. Kansainväliset esimerkit osoittavat, että samalla kertaa on mahdollista sekä hyödyntää massadatan koko potentiaalia että noudattaa tietosuojalainsäädäntöä ottamalla käyttöön uudenlaisia ratkaisuja, jotka vaikuttavat toimintatapoihin ja jopa julkisen sektorin toimijoiden rooleihin.

Toinen tapaustutkimuksista ilmi käynyt trendi ovat uudet työkalut ja standardit, jotka auttavat ymmärtämään hyvinkin abstraktia ja monimutkaista tietosuojalainsäädäntöä, kuten hollantilainen henkilötietoja sisältävän datan uudelleenkäytön työkalu osoittaa.

Kolmanneksi tapaustutkimukset osoittavat, että tietosuoja ei ole absoluuttinen oikeus. Joissakin tilanteissa periaatepunninta tietosuojan ja julkisten intressien välillä on tarpeen ottaen huomioon yhteiskunnalle syntyvät merkittävät hyödyt. Tietosuoja on kuitenkin vain yksi oikeudellinen osa-alue, jota muutos koskettaa. Olisi syytä huolellisesti arvioida yleisemminkin kansainvälisten innovatiivisten massadatan hyödyntämistä tukevien ratkaisujen soveltuvuutta ja kehitysmahdollisuuksia Suomessa.

Uusien työkalujen kehittäminen luo mahdollisuuksia tehostaa julkisen sektorin toimintaa, parantaa sen palveluita ja lisätä päätöksenteon ja resurssien allokoinnin avoimuutta sekä edistää myös massadatan hyödyntämistä (esim. datan uudelleen käyttöä eri tarkoituksiin) sekä yksityisellä että julkisella sektorilla.

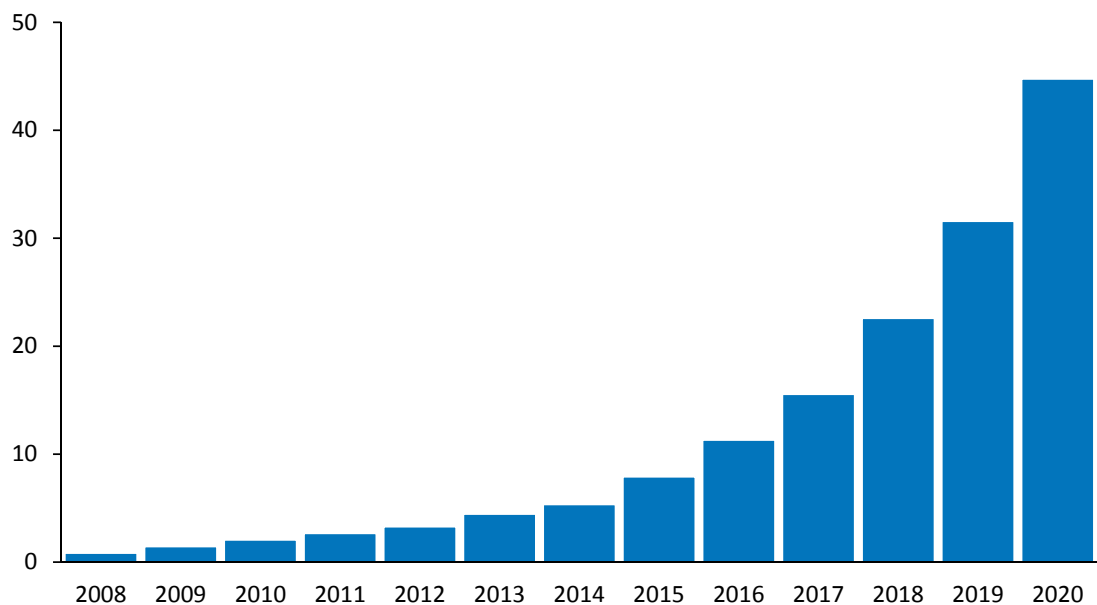
1 Johdanto

Teknologisen kehityksen myötä talouden eri sektoreilla pystytään hyödyntämään uudella tavalla sekä vanhaa järjestymätöntä dataa että muun muassa Internet-palveluista, kuluttajapalveluiden aika- ja paikakatiedoista, teknisten laitteiden toiminnasta, sensoriverkoista ja koneiden välisestä kommunikaatiosta (machine-to-machine communications) syntyvää uutta massadataa. Vielä vuonna 2000 vain noin neljännes maailman tietovarannoista oli digitaalisessa muodossa, kun vuonna 2013 yli 98 prosenttia siitä oli varastoitu digitaalisessa muodossa (Mayer-Schönberger ja Cukier, 2013). Datavarannot ovat pääosin organisaatioiden hallinnassa, kun taas suurin osa uudesta datasta syntyy yksityisten henkilöiden tuottamana. On arvioitu, että kaikesta datasta 80 prosenttia hallinnoivat yritykset ja siitä tuottavat 70 prosenttia yksittäiset henkilöt.¹

Datan määrässä odotetaan keskimäärin 40 prosentin vuosikasvua, ja sekä massadatateknologioiden ja -palveluiden markkinoiden että datatalouden kokonaisuudessaan odotetaan kasvavan voimakkaasti. Maailman massadatateknologioiden ja -palveluiden markkinoiden on arvioitu kasvavan vuosittain keskimäärin noin 23 prosenttia ja markkinoiden koon ylittävän 48 miljardia USD vuonna 2019². EU-alueen datatalouden arvoksi vuonna 2014 arvioitiin liki 250 miljardia euroa, eli 1,8 prosenttia alueen BKT:sta (IDC, 2015). Suomessa datatalouden osuuden arvioitiin kattavan noin 2,5 prosenttia BKT:sta.

Kansainvälisten tutkimusten mukaan massadatan suurimmat mahdollisuudet ja hyödyt syntyvät tehokkuuden tuomista kustannussäästöistä ja analytiikan mahdollistamasta nopeammasta ja täsmäl-

Kuvio 1 Dataperusteisen liiketoiminnan raaka-aineen määrä kasvaa nopeasti. Hyödynnettävissä olevan datan määrä maailmassa, zettabyteä (zb).



Lähde: Oracle.

¹ Computer Sciences Corporation, Big Data Universe Beginning to Explode. http://www.csc.com/insights/flxwd/78931-big_data_universe_beginning_to_explode

² <http://www.cio.com/article/3004512/big-data/idc-predicts-big-data-spending-to-reach-48-6-billion-in-2019.html>

lusemmästä päätöksenteosta sekä paremmista tuotteista ja palveluista (Davenport, 2014). Suomessa massadatan hyödyntämisestä ja liiketoimintapotentiaalista eri toimialoilla ja hyödyntämispotentiaalista julkisen sektorin palveluissa on tällä hetkellä varsin vähän tutkimukseen perustuvaa tietoa. Tämän tutkimushankkeen tarkoituksena oli tuottaa tietoa, jota voidaan käyttää digitaalisen kasvuympäristön rakentamiseen tähtäävän hallituksen kärkihankkeen toimissa. Kärkihankkeen tavoitteena on lisätä massadatan hyödyntämistä ja käynnistää omadata-kokeiluja. Lisäksi tutkimushanke tukee EU:n henkilötietoasetuksen pohjalta uudistumassa olevan tietosuojalainsäädännön kansallista valmistelua tunnistamalla tiedon käyttöön liittyviä oikeudellisia reunaehdoja. Tässä raportissa arvioidaan massadatan käyttöä ja liiketoimintapotentiaalia eri toimialoilla sekä massa- ja omadatan sovelluspotentiaalia julkisen sektorin palveluissa ja julkisen sektorin palveluihin liittyviä oikeudellisia reunaehdoja.

Luvussa 2 valotetaan marraskuussa 2015 toteutetulla kyselytutkimuksella kerätyn aineiston avulla massadatan hyödyntämistä ja liiketoimintapotentiaalia suomalaisissa yrityksissä ja toimialoilla. Luku 3 ja 4 käsittelevät massadataekosysteemiä ja massadataan perustuvien (pääasiassa julkisten) palveluiden kehittämiseen liittyviä ongelmakohtia ja pullonkauloja Suomessa ja niiden ratkaisemiseksi tarvittavia keinoja. Aineistona on käytetty kirjallisuuskatsausta, haastatteluja ja työpajoja. Luku 5 esittelee kansainvälisiä innovatiivisia esimerkkejä ratkaisuksista, joilla massatietoihin perustuvien yksityisten ja julkisten palveluiden kehittämistä voidaan edistää niin, että samalla turvataan riittävä tietosuojaa. Luvussa 6 esitetään lyhyesti johtopäätökset ja suosituksia toimenpiteistä, joilla voidaan edistää massadatan hyödyntämistä sen tarjoaman liiketoimintapotentiaalın realisoimiseksi ja julkisten palveluiden tehostamiseksi.

Keskeisiä käsitteitä

Massadata (suuraineisto) tai **big data** tarkoittaa dataa, jota on paljon, jota tulee nopeasti lisää ja joka on muodoltaan vaihtelevaa ja järjestymätöntä. Laitteilta ja sovelluksilta vaaditaan suurta tallennuskapasiteettia ja suorituskykyä massadatan käsittelyssä.

Internet of Things (IoT). Asioiden/esineiden internet (joissakin yhteyksissä myös teollinen internet) tarkoittaa Internetin laajentumista fyysiseen ympäristöön: esineisiin, rakennuksiin, laitteisiin ja koneisiin, joita voidaan ohjata, sensoroida ja mitata. IoT tuottaa erittäin suuria massadata-aineistoja, ja datamäärän pitäminen hallittuna on yksi IoT:n haasteista.

Avoim data tarkoittaa (pääasiassa) julkisen sektorin keräämää ja tuottamaa dataa, joka on kansalaisten ja kehittäjien käytettävissä konekielisenä sekä vapaasti ja veloituksettomasti käytettävissä, uudelleenkäytettävissä ja muokattavissa. Avoimesta datasta osa on massadataa.

Omadata (MyData). Omadatalle tarkoitetaan yhtäältä eri lähteissä sijaitsevaa yksilöön liittyvää tietoa, joka on konekielisesti saatavilla ja jonka käyttö on yksilön hallittavissa ja toisaalta myös tällaisen tiedon hallinnointiperiaatteita tai hallinnointia^a.

Älykkäät kaupungit (Smart Cities). Älykkäällä kaupungilla tarkoitetaan kaupunkia, joka käyttää laajamittaisesti tieto- ja viestintäteknologiaa kaikilla kaupungin toimialoilla palvelujen tuottamiseen tehokkaasti ja ympäristöystävällisesti. Älykkäät kaupungit hyödyntävät IoT:tä ja tuottavat samalla massadataa, joka on myös avointa dataa.

^a Poikola, ym. (2014). Big Data -strategiassa (LVM, 2014, s. 53) omadata määritellään seuraavasti: "Omadata (mydata/midata) viittaa toimintakulttuuriin, jossa organisaatio palauttaa keräämäänsä yksilökohtaista tietoa ihmiselle itselleen. Tämän jälkeen ihminen voi hyödyntää näin saamaansa tietoa suoraan tai jakaa sen edelleen haluamallaan tavalla. Jaettu tieto voi olla analysoimaton raakadata tai se voidaan jalostaa esimerkiksi visualisoimalla muotoon, jossa ihminen itse ja muut voivat sitä hyödyntää. Kansalaisten tulee voida nykyistä paremmin vaikuttaa siihen, mihin ja miten heitä koskevia tietoja käytetään. Hänen tulisi halutessaan voida hallita omia tietojaan ja voida luovuttaa kerättyä tietoa myös eteenpäin."

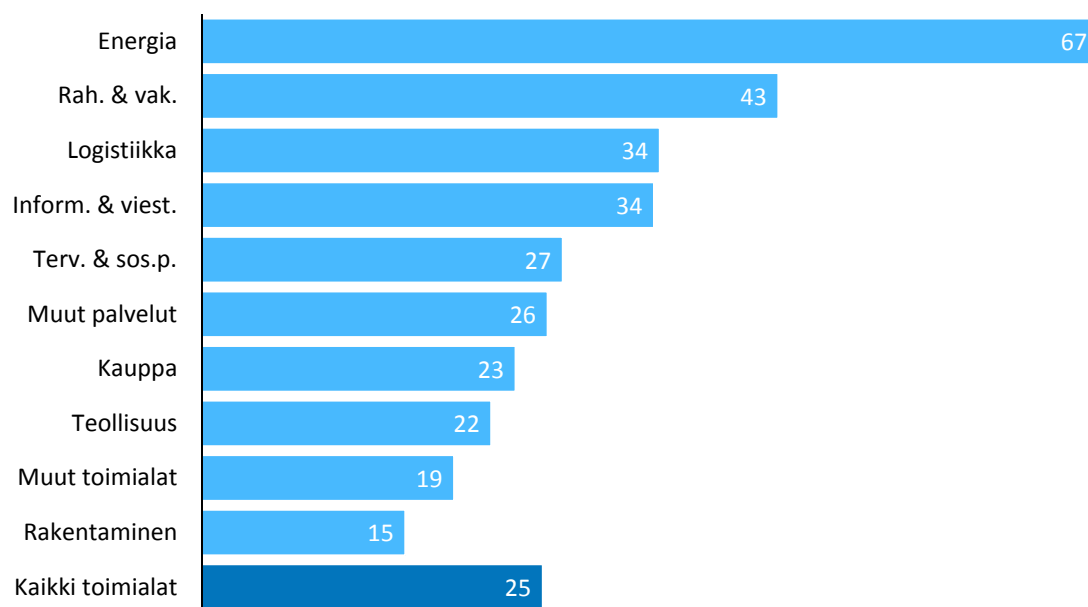
2 Massadatan käyttö ja liiketoimintapotentiaali suomalaisissa yrityksissä

Tämä luku valottaa suomalaisyritysten massadatan käyttöä ETLAn marraskuussa 2015 toteuttamaan yritys-kyselyyn perustuen. Massadatan hyödyntämisen eturintaman yritykset odottavat merkittävää massadatan strategiseen hyödyntämiseen perustuvaa kasvua. Isossa joukossa suomalaisia yrityksiä ei massadatan käytön liiketoimintapotentiaalia kuitenkaan vielä tällä hetkellä ymmärretä. Myös massadatan käsittelyyn ja analyysiin liittyvä tietotaidon puute rajoittaa tai estää massadatan hyödyntämistä merkittävästi.

2.1 Suuret yritykset massadatan käytön eturintamassa

Kyselyn kohdejoukkona oli 4 243 yritystä, joiden liiketoimintaan liittyi jotenkin digitalisaatio keväällä 2015 toteutetun yritys-kyselyn perusteella.³ Lisäksi kysely lähetettiin kaikille vähintään 50 henkeä työllistävälle yrityksille riippumatta siitä, olivatko ne vastanneet kevään kyselyyn. Linkki nettikyselyyn lähetettiin yhteensä 5 818 yritykselle, joista kyselyyn vastasi 1 189 yritystä (eli noin 20 prosenttia).

Kuvio 2 Massadatan käyttö yrityksissä toimialoittain, % yrityksistä



Lähde: ETLAn kysely, marraskuu 2015.

³ Yrityksiltä kysyttiin kevään 2015 kyselyssä useita kysymyksiä digitalisaation roolista niiden liiketoiminnassa, kuten ovatko yrityksen tuotteet digitaalisia, myykö se tuotteita verkkokaupassa ja panostaako yritys massadataan/pilvipalveluihin/robotiikkaan/avoimeen dataan/sosiaaliseen mediaan.

Yrityskyselyn alussa esitettiin massadatan määritelmä⁴ täsmentämään lomakkeessa käytettyä massadata-käsitettä. Lähes neljäsosa kyselyyn vastanneista yrityksistä keräsi massadataa tai käytti sitä liiketoiminnassaan. Vastaajien jakauma toimialoittain löytyy liitteestä 1. Kyselyssä ”Muut palvelut” -luokkaan luettiin ammatillinen, tieteellinen ja tekninen toiminta (sisältäen muun muassa arkkitehti- ja insinööri-palvelut ja mainostoiminnan ja markkinatutkimuksen), hallinto- ja tukipalvelut sekä koulutus. ”Muut toimialat” -luokkaan sisältyivät taulukossa mainittuihin luokkiin kuulumattomat yritykset.

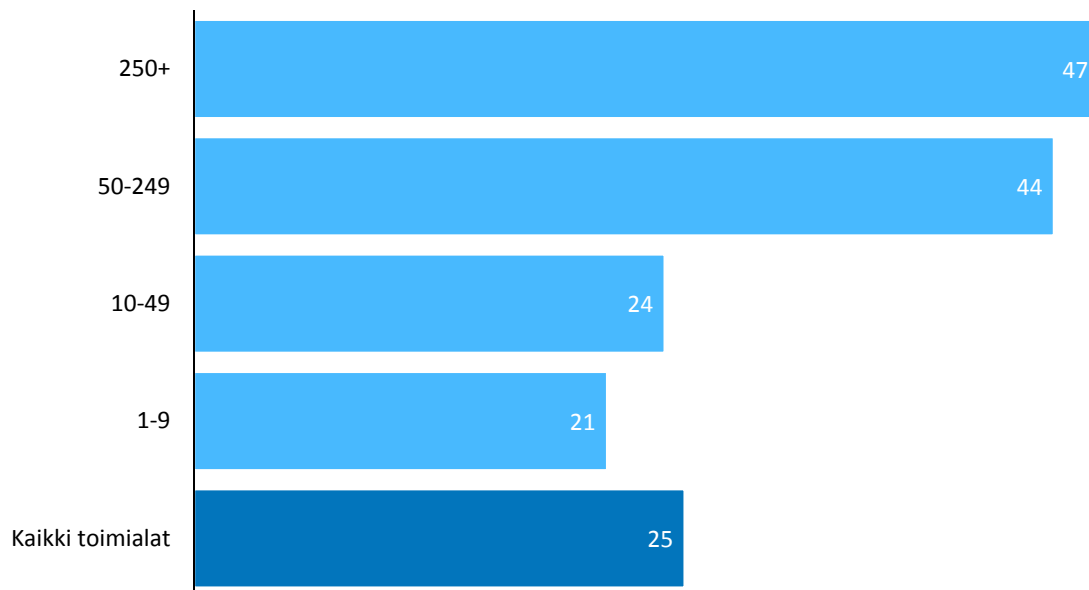
The U.S. Bureau of Economic Analysis ja McKinsey Global Institute arvioivat, että parhaat mahdollisuudet ja kyky luoda arvoa massadastasta on rahoitus- ja vakuutustoimintaa harjoittavilla yrityksillä (Gutierrez, 2014; Wolkowitz ja Parker, 2015). Näiden yritysten voidaan olettaa olevan massadatan hyödyntämisen etujoukoissa. Kyselyyn vastanneista rahoitus- ja vakuutustoiminnan yrityksistä 43 prosenttia oli massadatan käyttäjiä. Tosin toimialan yrityksiä oli vastanneiden joukossa vain 14. Vastauksen perusteella ei voida täten tehdä pidemmälle meneviä johtopäätöksiä rahoitus- ja vakuutustoiminnan yritysten massadatan hyödyntämisestä Suomessa.

Myös energia-alan yritykset ovat kansainvälisten arvioiden mukaan massadatan käyttöönoton eturintamassa. Vastanneista sähkö-, kaasu- ja lämpöhuoltoa harjoittavista 15 yrityksestä 67 prosenttia hyödynsikin massadataa liiketoiminnassaan. Massadatan käytöllä on nähtävissä selkeää potentiaalia myös logistiikka-alalla. Kuljetus- ja varastointialan yrityksistä yli kolmannes raportoi yrityksensä hyödyntävän massadataa liiketoiminnassaan. Massadatan hyödyntäminen oli yhtä yleistä informaatio- ja viestintäsektorin yritysten joukossa.

Massadatan hyödyntäminen oli yleisempää keskisuurissa ja suurissa (eli yli 50 henkeä työllistävissä) yrityksissä. Alle 10 henkeä työllistävistä yrityksistä noin viidesosa raportoi keräävänsä tai käyttävänsä massadataa, kun suurten yritysten joukosta massadatan käyttäjiä oli liki puolet.

Edellä esitetyt luvut massadatan hyödyntäjien osuudesta eivät edusta koko yrityspopulaatiota. Pk-yritysten osalta kyselytutkimuksemme rajattiin liiketoiminnassaan digitalisaatiota hyödyntäviin yrityksiin. Näiden joukossa massadatan käyttö on laajemmin levinnyttä kuin kaikissa pk-yrityksissä. Myös kyselyn aihe ”datan käyttö liiketoiminnassa” lienee rajannut vastaajajoukkoa niin, että vastaajien joukossa massadatan käyttö on todennäköisempää kuin koko yrityspopulaatiossa. Suomen yrittäjien PK-yritysbaremetriin⁵ vastanneista yrityksistä kolme prosenttia raportoi käyttävänsä big dataa. Kan-

Kuvio 3 Massadatan käyttö yrityksissä kokoluokittain (henkilöstömäärän mukaan), % yrityksistä



Lähde: ETLAn kysely, marraskuu 2015.

⁴ Katso tämän raportin luku 1.

⁵ PK-yritysbaremetri, kevät 2016, http://www.yrittajat.fi/File/51e24e66-00a0-4e36-8425-b8d96eed0f1c/baro_kevat2016_kalvot.pdf

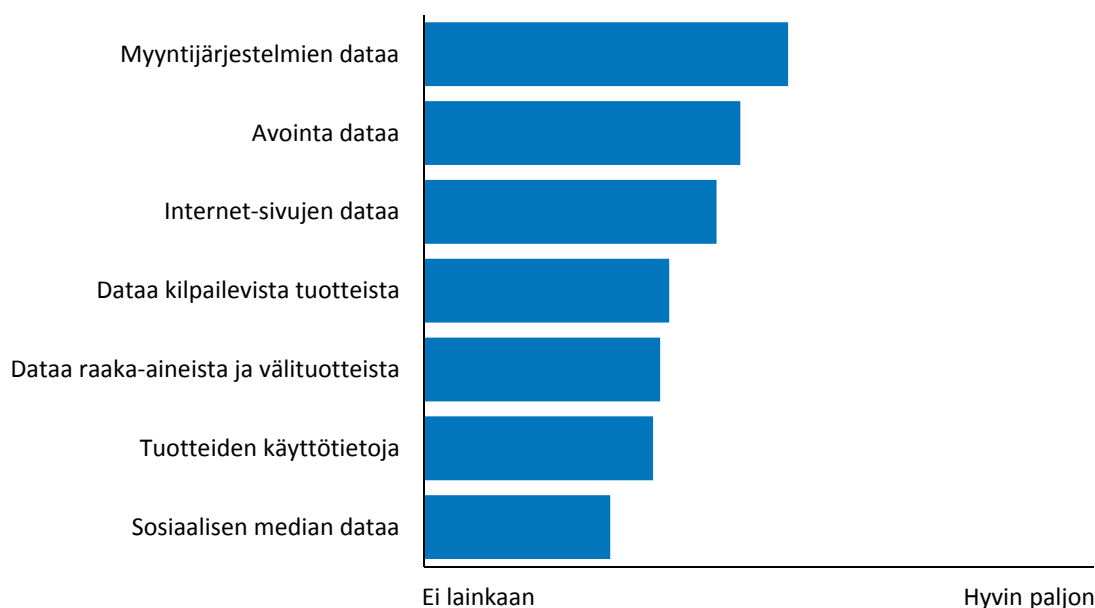
sainvälisissä raporteissa esitetyt arviot massadatan käytön leviämisestä viittaavat myös massadatan hyödyntämisen olevan suhteellisen varhaisessa vaiheessa kaikkien yritysten joukossa. Esimerkiksi e-skills UK -raportissa arvioitiin, että Isossa-Britanniassa isoista yrityksistä reilu neljännes olisi massadatan käyttäjiä vuonna 2015⁶, kun kyselyymme vastanneista isoista yrityksistä liki puolet käytti massadataa.

Tärkeimpiä massadatana hyödynnettyjä tietoja olivat yrityksen myyntijärjestelmistä (esim. kan- ta-asiakasjärjestelmän kautta) ja yrityksen www-sivujen kautta kerätty data. Vastaajista 61 prosenttia kertoi yrityksessään hyödynnettävän myyntijärjestelmistä kerättyä dataa ja 57 prosenttia www-sivujen kautta kerättyä dataa. Myyntijärjestelmien asiakastietoja hyödynsi hyvin paljon liiketoiminnassaan massadatana 31 prosenttia vastaajista, kun www-sivujen kautta kerättyä dataa hyvin paljon hyödyn- täviä oli 18 prosenttia. Avoimeen kysymykseen muista yrityksen liiketoiminnassa hyödyntämistä da- toista mainittiin muun muassa tutkimusdata kuluttajien käyttäytymisestä, asenteista ja mediankäytöstä, GPS:llä liikkumisesta kerrytetty paikkatiedot ja suurikokoiset kuva- ja videotiedostot.

Noin 35 prosenttia yrityksistä kertoi ostavansa massadataa muilta yrityksiltä tai organisaatioilta. Paljon tai hyvin paljon massadataa osti 12 prosenttia massadatan hyödyntäjistä. Massadataa myi vain 15 prosenttia yrityksistä, ja kysymykseen vastanneista yrityksistä seitsemän prosenttia paljon tai hyvin paljon. Suurempi osa, eli 38 prosenttia, kertoi jakavansa dataa muiden yritysten tai organisaatioiden kanssa muilla tavoin. Noin 11 prosenttia yrityksistä kertoi jakavansa massadataa muiden kanssa pal- jon tai hyvin paljon. Tilastokeskuksen ennakkotietojen mukaan yritysten innovaatiotoiminta -kyselyyn vastanneista yrityksistä massadataa oli vuosina 2012–2014 ostanut noin 22 prosenttia ja myynyt noin 15 prosenttia vastaajista.⁷

Yleisimmät toiminnot, joissa massadataa hyödynnettiin, olivat päätöksenteko (62 prosenttia yri- tyksistä), markkinointi ja myynti (59 prosenttia yrityksistä), tavaroiden ja/tai palveluiden tuottaminen (56 prosenttia yrityksistä) ja asiakas- ja markkina-analyysi (55 prosenttia yrityksistä). Massadatan käy- töstä saadaan hieman erilainen kuva, jos katsotaan massadatan merkittävyyttä eri toiminnoissa. Mas- sadataa paljon tai hyvin paljon hyödynsi päätöksenteossa noin 40 prosenttia vastanneista ja asiakas- ja markkina-analyysissa jonkin verran alle 40 prosenttia yrityksistä. Kolmanneksi tärkeimmäksi nousi

Kuvio 4 Millaista massadataa hyödynnetään liiketoiminnassa?



Lähde: ETLAn kysely, marraskuu 2015.

⁶ e-skills UK (2014). Big Data Analytics: Adoption and Employment Trends, 2012–2017. <http://www.sas.com/offices/europe/uk/downloads/bigdata/eskills/eskills.pdf>

⁷ Innovaatiotoiminta 2014, Tilastokeskus. Sähköpostitse saadut ennakkotiedot.

Kuvio 5 Massadatan hyödyntäminen yrityksen toiminnoissa. Niiden yritysten %-osuus, jotka hyödyntävät massadataa paljon tai hyvin paljon ko. toiminnossa



Lähde: ETLAn kysely, marraskuu 2015.

massadatan hyödyntäminen uusien tuotteiden kehittämisessä: noin 37 prosenttia yrityksistä ilmoitti käyttävänsä massadataa paljon tai hyvin paljon uusien tuotteiden kehittämisessä. Tilastokeskuksen yritysten innovaatiotoiminta -kyselyn ennakkotietojen mukaan vuosina 2012–2014 massadataa käytti uusien tuotteiden kehittämisessä noin 27 prosenttia yrityksistä, mutta vain vajaalle 10 prosentille massadatan merkitys uusien tuotteiden kehittämisessä oli kohtalainen tai suuri.⁸ Dataa käytti kyseisenä ajanjaksona tutkimus- ja kehittämistoiminnassa noin kolmannes vastaajayrityksistä.

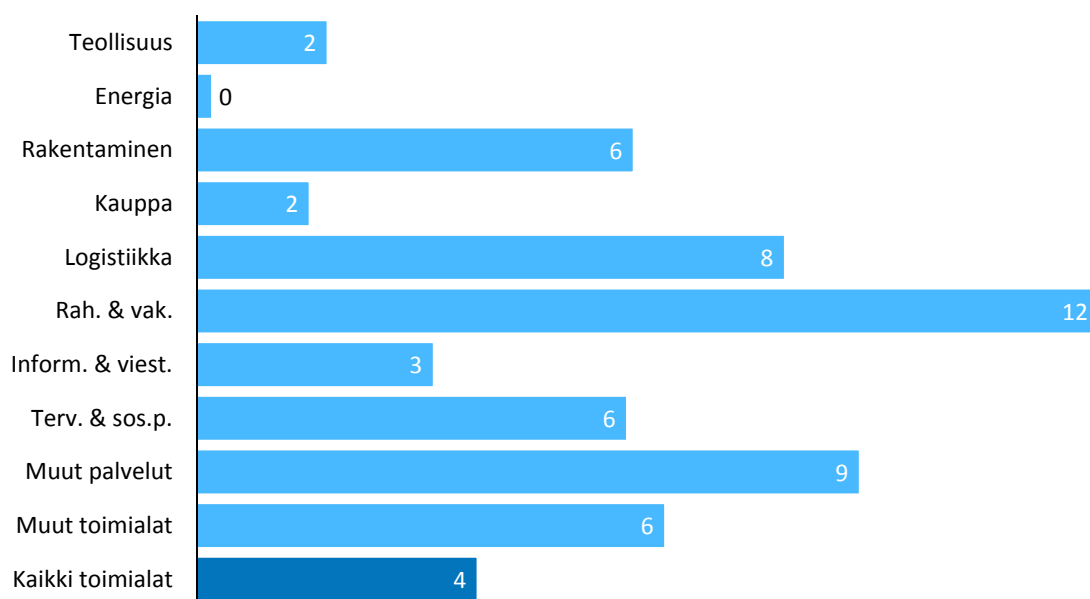
Kartoitimme kyselyssämme myös yrityksissä tällä hetkellä massadataan liittyvissä tehtävissä toimivien henkilöiden määrää kysymyksellä: ”Montako henkilöä yrityksessänne työskentelee päätoimisesti tiedonlouhintaan, massadatan analysointiin ja/tai mallintamiseen liittyvissä tehtävissä?”. Noin kaksi kolmasosaa massadatan hyödyntäjistä vastasi tähän kysymykseen. Vastaajayrityksissä työskenteli keskimäärin alle kolme henkilöä päätoimisesti massadataan liittyvissä tehtävissä, kun mediaani oli yksi henkilö. Noin 30 prosentissa massadataa hyödyntävistä vastaajayrityksistä ei ollut yhtään päätoimista massadataan liittyvissä tehtävissä toimivaa henkilöä.

Keskimäärin neljä prosenttia massadataa hyödyntävien vastaajayritysten työntekijöistä toimi päätoimisesti massadataan liittyvissä tehtävissä. Suurissa yrityksissä massadataan liittyvissä tehtävissä toimi noin prosentti ja pienissä, alle 10 henkeä työllistävissä yrityksissä noin puolet työntekijöistä. Toimialoista rahoitus- ja vakuutus työllisti eniten (eli noin 12 prosenttia) massadataan liittyvissä tehtävissä toimivia suhteutettuna yrityksen työntekijöiden lukumäärään. Yrityskyselystä saatu keskiarvo massadataan liittyvien työntekijöiden suhteellisesta osuudesta on hyvin lähellä IDC:n arviota tietotyöntekijöiden (data workers) suhteellisesta työvoimaosuudesta: IDC arvioi, että Suomessa tietotyöntekijöiden osuus kaikista työllisistä oli vuonna 2014 noin 3,6 prosenttia (IDC, 2015).

Kysymykseen massadataan liittyvissä tehtävissä työskentelevien henkilöiden määrästä on vastannut valikoitunut joukko yrityksiä, jotka hyödynsivät kyselyn toteutushetkellä massadataa. Täten vastaukset massadatan parissa työskentelevien työntekijöiden määristä heijastanevat massadatan hyödyntämisen edelläkävijäyritysten tilannetta. OECD:n (2015) raportissa arvioidaan data-asiantuntijoiden määrää ISCO-08 (2008 International Standard Classification of Occupations) -luokitusta käyttäen laskien matematiikan ja tilastotieteen asiantuntijoiden (luokka 212) ja tietokantojen, -verkkojen ja jär-

⁸ Innovaatiotoiminta 2014, Tilastokeskus. Sähköpostitse saadut ennakkotiedot.

Kuvio 6 Päätoimisesti massadataan liittyvissä tehtävissä toimivien henkilöiden osuus yritysten koko henkilöstömäärästä, %



Lähde: ETLAn kysely, marraskuu 2015.

Lähes 50 % massadatan käyttäjistä ei osannut arvioida massadataan liittyvissä tehtävissä työskentelevien henkilöiden lukumäärää; kuviossa vain yritykset, joilta arvio saatiin.

jestelmäohjelmistojen erityisasiantuntijoiden (luokka 252) suhteellisia osuuksia työvoimasta. Tämän suppeamman luokittelun mukaan data-asiantuntijoiden osuus kaikista työntekijöistä oli Suomessa vuonna 2013 noin 0,5 prosenttia.

Evans Data Corporationin (2015)⁹ 444 ohjelmoijan haastatteluun perustuva tutkimus heijastelee myös massadatasovellusten kehittämisen laajuutta eri toimialoilla. Kolmannes haastatelluista massadata- ja analyyttikkasovelluksia kehittävästä ohjelmoijista työskenteli ohjelmistotalalla. Muita toimialoja, joilla massadataohjelmoijien osuus oli verrattain suuri, olivat rahoitussektori, teollisuus ja kaupan ala.

2.2 Massadatan liiketoimintapotentiaali: parempia päätöksiä, tuotteita ja asiakaskokemuksia

IDC:n (2015) esittämän arvion mukaan datatuotteiden ja palveluiden myynti oli vuonna 2014 Suomessa yli 800 miljoonaa euroa ja EU-alueella 51 miljardia euroa. Massadateknologioiden ja -palveluiden markkinoiden odotetaan kasvavan voimakkaasti. IDC¹⁰ arvioi maailman massadateknologioiden ja -palveluiden markkinoiden vuosittaiseksi kasvuksi noin 23 prosenttia ja markkinoiden kooksi yli 48 miljardia USD vuonna 2019. Merkittävimiksi massadataan investoiviksi toimialoiksi arvioidaan teollisuutta ja rahoitusalan yrityksiä. Tämä luo odotuksia massadatan liiketoimintapotentiaalin realisoinnista sen hyödyntäjäyrityksissä.

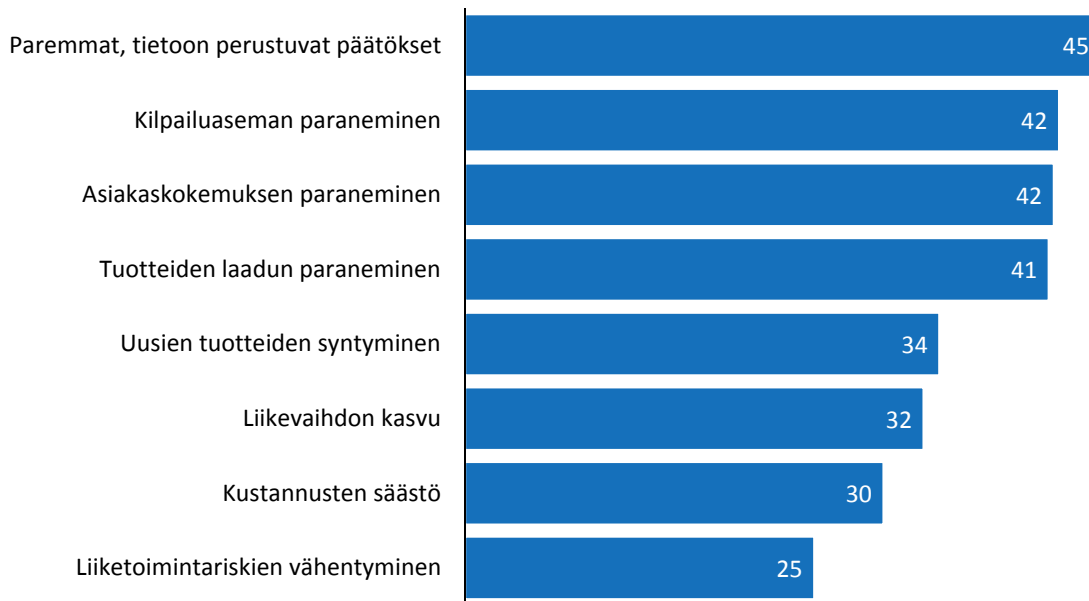
Massadatan hyödyntämisen tuoman lisäarvon mittaaminen on vaikeaa, ja sen liiketoimintapotentiaalista onkin esitetty suuruusluokaltaan hyvin vaihtelevia arvioita. IDC:n (2015) raportissa arvioidaan datatalouden osuudeksi Suomessa noin 2,5 prosenttia BKT:sta. Koko EU-alueella datatalouden arvoksi on laskettu vuonna 2014 lähes 250 miljardia euroa (eli 1,8 prosenttia EU-alueen BKT:sta).

Kyselyssämme yrityksiltä kysyttiin, missä määrin ne arvioivat massadatan hyödyntämisen vaikuttavan yrityksessään niiden liiketoimintaan nyt tai lähitulevaisuudessa. Yleisimmin (eli noin 45 prosenttia yrityksistä) vastaajat arvioivat massadatan käytön tuottavan parempia, tietoon perustuvia pää-

⁹ Evans Data Corporation (2015). Big data and advanced analytics survey. Vol 1. <http://www.forbes.com/sites/louiscolumnbus/2015/10/11/2015-big-data-market-update/#394f2a7f5170>

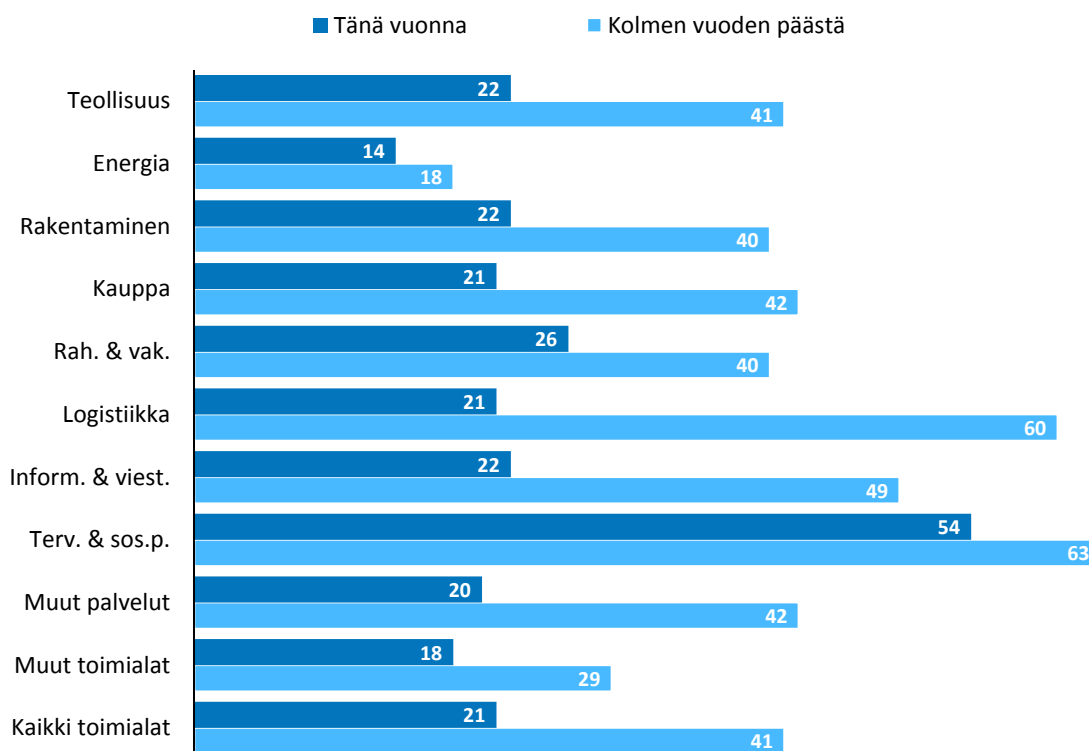
¹⁰ <http://www.cio.com/article/3004512/big-data/idc-predicts-big-data-spending-to-reach-48-6-billion-in-2019.html>

Kuvio 7 Massadatan käytön liiketoimintapotentiali nyt tai lähitulevaisuudessa.
%-osuus yrityksistä, jotka arvioivat liiketoimintapotentialin olevan melko tai hyvin suuri.



Lähde: ETLAn kysely, marraskuu 2015.

Kuvio 8 Massadatan käytön osuus liikevaihdosta eturintaman hyödyntäjäyrityksissä, mediaani-%



Lähde: ETLAn kysely, marraskuu 2015.

Noin 30 % massadatan käyttäjistä ei osannut arvioida liikevaihto-osuutta; kuviossa vain yritykset, joilta arvio saatiin.

töksiä yrityksessään. Yli 40 prosenttia vastaajista arvioi massadatan hyödyntämisen myös parantavan yrityksen kilpailuasemaa, tuotteiden laatua ja asiakaskokemusta paljon tai hyvin paljon.

Liitteessä 1 on esitetty massadatan käytön odotetut liiketoimintavaikutukset toimialoittain. Kuviosta nähdään, että rahoitus- ja vakuutustoimintaa harjoittavissa yrityksissä odotetaan muita useammin massadatan käytön kasvattavan yrityksen liikevaihtoa, parantavan tuotteiden laatua, säästävän kustannuksia, vähentävän liiketoimintariskejä sekä johtavan parempiin, tietoon perustuviin päätöksiin.

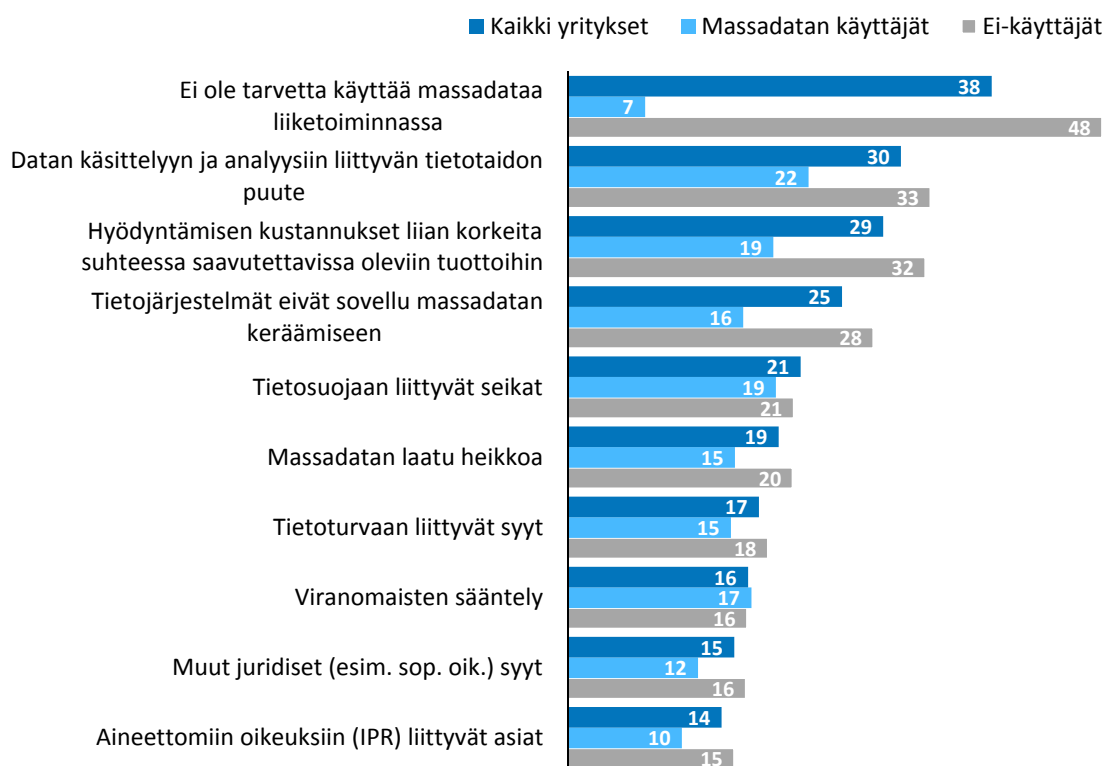
Lähes kolmannes massadataa käyttävistä yrityksistä arvioi massadatan hyödyntämisen vaikuttavan paljon tai hyvin paljon niiden liikevaihdon kasvuun nyt tai lähitulevaisuudessa. Noin 30 prosenttia näistä yrityksistä ei osannut vastata tarkempaan kysymykseen siitä, mikä osuus yrityksen liikevaihdosta perustuu massadatan hyödyntämiselle tänä vuonna tai kolmen vuoden päästä. Kysymykseen vastanneiden hyödyntäjäyritysten tyypillinen arvio oli, että massadatan hyödyntämisen osuus niiden liikevaihdosta on tänä vuonna noin viidesosa ja kolmen vuoden päästä 41 prosenttia. Suurinta kasvua odotettiin terveys- ja sosiaalipalveluita tarjoavissa sekä logistiikka-alan yrityksissä.

Vastauksia ja kuviota 8 tulkittaessa on tärkeää huomata, etteivät kuvion toimialoittaiset prosentit edusta tyypillisiä toimialan yrityksiä vaan valikoitunutta joukkoa massadatan hyödyntäjiä. Toisin sanoen, liikevaihtovaikutukset ovat arvioita massadatan hyödyntämisen vaikutuksista niissä yrityksissä, jotka ovat massadatan käytön eturintamassa.

2.3 Massadatan hyödyntämisen rajoitteet: tarpeita ei nähdä, tietotaitoa puuttuu

Kyselylomake sisälsi kysymyksen tekijöistä, jotka rajoittavat tai estävät massadatan hyödyntämistä yrityksen liiketoiminnassa. Tämä kysymys kysyttiin myös niiltä yrityksiltä, jotka eivät käytä massadataa. Merkittävimmäksi tekijäksi nousi se, ettei yrityksissä nähty tarvetta hyödyntää massadataa tai ymmärretty massadatan hyödyntämisen liiketoimintapotentiaalia. Lähes puolet niistä yrityksistä, jotka eivät käyttäneet massadataa, ei nähnyt tarvetta sen käytölle liiketoiminnassaan.

Kuvio 9 Massadatan käytön rajoitteet tai esteet. %-osuus yrityksistä, jotka kokevat rajoitteen tai esteen melko tai hyvin suurena.



Lähde: ETLAn kysely, marraskuu 2015.

Massadatan käsittelyyn ja analyysiin liittyvä tietotaidon puute rajoitti tai esti massadatan hyödyntämistä paljon tai hyvin paljon noin 30 prosentissa vastaajayrityksistä. Massadatan hyödyntäjistä noin 22 prosenttia piti tietotaidon puutetta rajoitteena. Liitteen 1 toimialakohtainen taulukko osoittaa, että erityisesti logistiikka-alan massadataa hyödyntävät yritykset pitivät tietotaidon puutetta ongelmana. Tämä saattaa osaltaan selittää havainnon, että suhteellisen harva yritys hyödyntää massadataa hyvin paljon tai paljon logistiikassa. Liki kolmannes massadatan ei-käyttäjistä piti myös merkittävänä massadatan hyödyntämisen rajoitteena tai esteenä saavutettavissa oleviin tuottoihin nähden liian korkeita kustannuksia. Tätä seikkaa piti esteenä tai rajoitteena alle viidesosa massadataa hyödyntävistä yrityksistä.

Viranomaisten sääntelyä, aineettomiin oikeuksiin (IPR) liittyviä asioita ja muihin juridisiin (esim. sopimusoikeudellisiin) kysymyksiin sekä tietoturvaan liittyviä syitä pidettiin vähiten massadatan hyödyntämistä rajoittavina tai estävinä tekijöinä sekä massadatan käyttäjien että ei-käyttäjien joukossa. Myös IDG Enterprisen (2015) raportoimat yrityskyselyiden tulokset viittaavat siihen, että yritysten luottamus tietoturvaratkaisuihin on vahvistunut. Vuonna 2014 IDG:n verkkokyselyyn vastanneista henkilöistä 44 prosenttia uskoi yrityksen olemassa olevien tietoturvaratkaisujen ja -tuotteiden olevan riittäviä takaamaan tietoturvan yrityksen tietovarannoille, kun vastaava prosenttiluku vuonna 2015 oli 66.

2.4 Miten massadatan hyödyntäjät poikkeavat muista yrityksistä?

Kyselyaineiston perusteella tehtiin tilastollinen analyysi, jonka avulla haettiin vastausta siihen, miten massadataa hyödyntävät yritykset poikkeavat muista yrityksistä. Edellä esitetyt kuviot antavat suuntaa-aviivaa yritysten välisistä eroista, mutta eivät kerro sitä, mitkä erot massadatan hyödyntäjien ja muiden yritysten välillä ovat tilastollisesti merkittäviä.

Massadatan hyödyntäjien eroa muista yrityksistä selvitettiin estimoimalla probit-malli, jossa selitettävä muuttuja sai arvon 1, jos yritys hyödynsi massadataa ja arvon 0, jos se ei hyödyntänyt massadataa. Potentiaalisina selittävinä muuttujina olivat yrityksen ikä, koko, toimiala, maakunta sekä erilliset indikaattorimuuttujat sille, harjoittaako yritys digitaalisten ja ei-digitaalisten palveluiden ja/tai tuotteiden verkkokauppaa.

Massadatan hyödyntäjät eivät eronneet muista yrityksistä tilastollisesti merkittävästi iän suhteen. Sen sijaan yrityksen koolla oli merkitystä. Yrityksen todennäköisyys hyödyntää massadataa kasvoi koon myötä. Keskisuuret ja isot yritykset olivat selkeästi todennäköisemmin massadatan hyödyntäjiä kuin pienemmät yritykset. Myös 10–49 henkeä työllistävät yritykset hyödynsivät massadataa todennäköisemmin kuin alle 10 henkeä työllistävät yritykset.

Massadatan hyödyntämisen laajuudessa oli tilastollisesti merkittäviä toimialoittaisia eroja. Yritykset, jotka toimivat informaatio ja viestintä -toimialalla olivat muiden toimialojen yrityksiä todennäköisempiä massadatan käyttäjiä. Malli estimoitiin myös käyttämällä Eurostatin tilastoinneissa käyttämää toimialojen teknologialuokitusta, jossa on erillisiä luokkina tietointensiiviset palvelut, teollisuus ja otellaan niiden teknologiaintensiivisyyden (eli yrityksen t&k-menot suhteessa sen arvonlisään) perusteella korkeaan, keskikorkeaan, keskimatalaan ja matalaan ja lisäksi omana kategorianaan ovat ”muut alat”.¹¹ Teollisuuden keskimatalalla aloilla toimivat yritykset käyttivät massadataa muita yrityksiä selvästi harvemmin, kun taas tietointensiiviset palvelut -alaan lukeutuvat yritykset olivat muita todennäköisemmin massadatan hyödyntäjiä.

Massadatan hyödyntämisessä löytyi myös tilastollisesti merkittäviä alueellisia eroja. Estimoinneissa olivat verrokkina Uudenmaan yritykset. Suurimman osan 20 muussa maakunnassa sijaitsevan yrityksen todennäköisyys käyttää massadataa ei poikennut verrokkiryhmästä. Sen sijaan Keski-Suomen yritykset olivat tilastollisesti merkittävästi harvemmin massadatan hyödyntäjiä kuin muut yritykset. Pohjois-Pohjanmaalla toimivat yritykset käyttivät todennäköisemmin massadataa kuin muut yritykset.

Yrityksen verkkokauppatoiminta vaikutti myös likeisesti liittyvän massadatan hyödyntämiseen. Sekä digitaalisia että ei-digitaalisia tuotteita ja/tai palveluita Internetin kautta myyvät yritykset olivat muita yrityksiä todennäköisemmin myös massadatan hyödyntäjiä.

11 Tarkempaa tietoa luokituksesta löytyy Eurostatin sivuilta: http://ec.europa.eu/eurostat/cache/metadata/DE/htec_esms.htm

2.5 Päätelmät ja suositukset

Massadatan hyödyntämisen etujoukoissa oleville yrityksille massadata mahdollistaa jo nyt parempien, tietoon perustuvien päätösten tekemisen, uusien tuotteiden kehittämisen ja tuotteiden laadun parantamisen sekä parempien asiakaskokemusten tarjoamisen. Kyky hyödyntää massadataa nähdään yrityksen kilpailuasemaa vahvistavana.

Monilla sektoreilla massadatan hyödyntämisessä ollaan kuitenkin varsin varhaisessa vaiheessa, eikä yrityksissä ymmärretä täysin massadatan tarjoamia liiketoimintamahdollisuuksia. Yrityksissä ei myöskään ole riittävästi massadatan hyödyntämiseen liittyvää osaamista. Datan käsittelyyn ja analyysiin liittyvä tietotaidon puute on merkittävä massadatan hyödyntämisen este yrityksissä.

Massadatan tehokkaampi hyödyntäminen eri toimialoilla vaatii ymmärrystä sen tarjoamasta potentiaalista ja panostusta osaamiseen. Näiden tavoitteiden saavuttamiseksi olisi ensinnäkin suunnattava resursseja osaamisen riittävyyden varmistamiseksi datan käsittelyyn, hallintaan ja analytiikan koulutukseen. Teknisten alojen lisäksi on syytä korostaa myös datatalouden yhteiskunnallis-taloudellisen koulutuksen merkitystä. Toiseksi, tarvitaan monialaista massadatan hyödyntämistä edistävää tutkimus- ja innovaatiotoimintaa. Teknisen tutkimuksen lisäksi tarvitaan datan hyödyntämiseen liittyvien laajempien yhteiskunnallis-taloudellisten mekanismien ymmärtämistä. Tämä on yksi edellytys onnistuneelle datapolitiikalle. Myös datan hyödyntämiseen liittyvät palvelukehityskokeilut voivat toimia yrityksille merkittävinä oppimislustoina massadatan liiketoimintapotentiaalin valjastamiseksi.

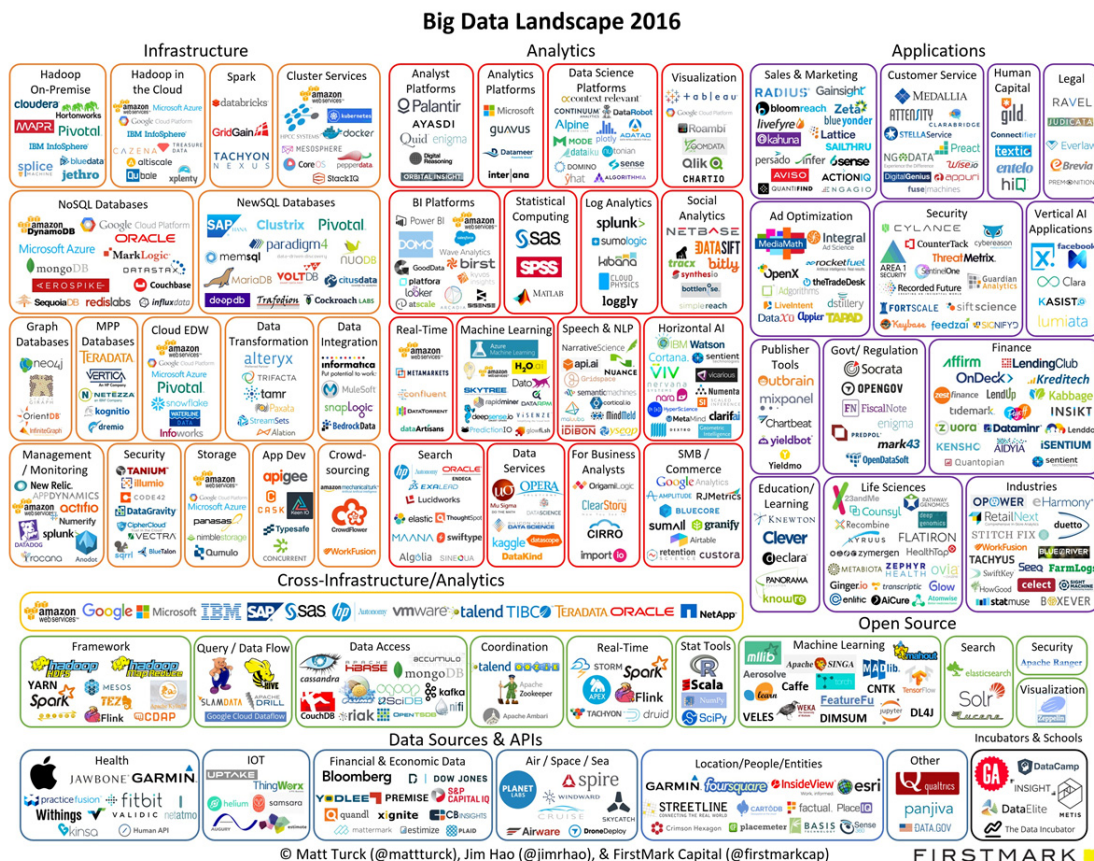
3 Massadata: nykykäytännöt ja tulevat kehityssuunnat

Tämän luvun tarkoituksena on lisätä ymmärrystä massadataan perustuvien yksityisten ja julkisten palveluiden kehittämispotentiaalista ja toimenpiteistä, joita tarvitaan näiden palveluiden luomisen edistämiseksi. Se kuvaa erityisesti massadatan hyödyntämisen nykykäytäntöjä, potentiaalia ja riskejä julkisella sektorilla. Omadata on kriittisessä asemassa massadatan hyödyntämisessä. Haasteiden ratkaisemiseen tarvitaan uusia toimintamalleja, kuten omadata -ekosysteemiä, missä painotus siirtyy lainsäädännöstä käyttäjien hallitsemiin datan hallinnan työkaluihin.

3.1 Talouden dataistuminen: Massadataan liittyvät käsitteet ja riskit

OECD:n (2015) Big Data -raportin mukaan kansantalouden dataistuminen koskettaa kaikkia aloja: terveyspalveluja, julkishallintoa, innovointia, koulutusta ja ympäristöä. Datan hyödyntämismahdollisuudet liittyvät mm. yksityisyydensuojaan, kuluttajapolitiikkaan, kilpailuun ja verotukseen. Jotta massadatan potentiaali voidaan valjastaa käyttöön, erityisesti pienten ja keskisuurten yritysten ja korkean

Kuvio 10 Massadata-maisema



Lähde: Turck ja Hao, 2016.

arvonlisän yritysten mukana olo datan avaamisessa, jalostamisessa ja käytössä palveluita rakennettaessa on tärkeää. Massadata on laaja käsite, joka liittyy läheisesti moniin muihin datan hyödyntämiseen perustuviin ilmiöihin. Tässä on avattu lyhyesti niistä tärkeimpien nykytilannetta.

Internet of Things (IoT) etenee tasaisesti kaikissa länsimaissa. Accenturen tuoreen tutkimuksen (Purdy ja Davarzani, 2015, s. 12) mukaan Suomen lähtökohtatilanne on hyvä: Suomi on IoT-valmius-tasoltaan kolmantena maailmassa USA:n ja Sveitsin jälkeen. Suomessa on osaamista ja perinteitä korkean teknologian soveltamisessa teolliseen ympäristöön, ja sekä isoja että pieniä alan yrityksiä. Molempia tarvitaan globaaleja arvoketjuja rakennettaessa. Suomessa on ympäristöystävällisten ja energiaa säästävien teknologioiden kehittämisessä edelläkävijäyrityksiä. Internet of Things etenee nopeasti, mutta alueella on vielä merkittäviä haasteita liittyen mm. teknologioiden täsmällisyyteen, tietoturvaan, järjestelmien kestävyys- ja huoltovarmuuteen. Yksi suurimpia haasteita on avoimien standardien puute, mikä rajoittaa skaalautuvuutta. Datamallit ovat kehittymättömiä ja lähiverkkoteknologiat suljettuja ratkaisuja. Yhtenäisiä standardeja ja rajapintoja kehitetään eri aloille ja alojen välille. Kaupungit ovat merkittävä IoT:n sovellusalue ja IoT on potentiaalisesti suurin kaupunkien massadatan lähde, koska lähes kaikki urbaanit järjestelmät voidaan liittää asioiden Internetiin.

Avoimen datan tilanne Suomessa on kansainvälisesti ottaen melko hyvä. Helsingin seutu on ollut alan pioneeri, ja on nykyisin maailman johtava alueellinen datan avaaja¹². Valtion Avoimen datan ohjelma tuottaa tuloksia; tärkeät sektoritutkimuslaitokset (esim. Maanmittauslaitos) ovat avanneet paljon dataa. Kokonaisuutena datan avaamisessa ollaan silti vasta alussa. Tärkein kehityskohde tulevina vuosina on riittävän suurten datamarkkinoiden rakentaminen. Tähän liittyen tarvitaan systemaattista datan avaamista, datan yhteiskäyttöisyyden edistämistä avoimilla standardeilla ja tietomallien harmonisoinnilla sekä parantamalla datan löydettävyyttä eri lähteistä, rajapintojen avaamista ja harmonisointia reaaliaikaisen datan avaamiseksi, datan hyödyntämistä mm. analytiikan ja tekoälyn avulla sekä kansainvälistä datayhteistyötä, jolla rakennetaan yhteistä eurooppalaista datan markkinapaikkaa.

Suomi on yksi harvoja maita, joissa hallitusohjelma painottaa omadata-periaatteita julkisia palveluita digitalisoitaessa. Merkittävä osa kaupunkien tuottamasta ja kokoamasta datasta sisältää henkilötietoja, jotka voisivat olla omadataa. Suomalainen alan tutkimus on korkeatasoista ja kansainvälisesti hyvin verkottunutta. Pohjoismaisena luottamusyhteiskuntana Suomella on hyvät mahdollisuudet profiloitua yhdeksi omadata-kehityksen kansainvälisistä osaamiskeskuksista. Toimivat omadata-ratkaisut (esim. koskien luottamuksellisia tietoja sisältäviä terveystietoja) tulevat olemaan ratkaisevassa asemassa digitalisaatiokehityksen jatkuessa kaikilla yhteiskunnan osa-alueilla. Omadata on kehittymässä oleva alue, johon liittyy merkittäviä haasteita niin lainsäädännön, tietotekniikan kuin liiketoimintamallienkin alueilla. Tarvittavia teknisiä ratkaisuja (identiteetin ja suostumusten hallinnan välineet) ei ole. Lainsäädäntö määrittelee tiedon kokoamisen ja käytön rajoituksia organisaatioille, mutta ei takaa ihmisille konekielistä pääsyä omaan dataansa. Omadata on pääosin suljettuna yritysten ja muiden organisaatioiden järjestelmiin, ja erityisesti ns. alustayhtiöt (Facebook ym.) ovat käyttöehdoissaan taanneet itselleen datan hyödyntämismahdollisuuden.

Älykkäät kaupungit (Smart Cities). Maailma kaupungistuu erittäin nopeasti, joten kestävämmille urbaaneille ratkaisuille on globaali tarve. Älykaupunkien ensimmäinen aalto ajoittuu 2000-luvun puoliväliin, jolloin käynnistettiin merkittäviä pilottihankkeita mm. Koreaan, Abu Dhabiin ja Kiinaan, ja suuryritykset kuten IBM ja Microsoft kehittivät omat smart city -tuotteensa. Näille oli tyypillistä vahva teknologiavetoisuus ja suljettu toimintamalli. Yksi merkittävimpiä älykaupunkiratkaisujen haasteista on markkinapaikan hajanaisuus. Kaupunkien haasteet ovat samankaltaisia (liikenne, ympäristö, vesi, turvallisuus, ym.), mutta taustat, syyt ja rakenteet erilaisia, jolloin samat ratkaisut eivät toimi eri kaupungeissa muokkaamatta. Lisäksi mahdollistavien teknologioiden standardit ovat kehittymättömiä, kaupunkien toimintamallit keskenään erilaisia ja osaamisesta pulaa. Suomalaisilla toimijoilla voi olla merkittäviä mahdollisuuksia tällä alueella. Monet tällaiset ratkaisut ovat varsin kehittyneitä kansainvälisessä vertailussa. Suomalaiset kaupungit ovat hallittavan kokoisia testausympäristöjä ja väestön teknologiaosaaminen korkealla tasolla.

12 Olli Sulopuisto, Citylab, 29.4.2014; <http://www.citylab.com/tech/2014/04/how-helsinki-mashed-open-data-regionalism/8994/>

Massadatan hyödyntämiseen liittyy myös merkittäviä riskejä:

- i) **Henkilötietoriskit.** Suuri massadatan riski liittyy henkilötietojen suojaamiseen. Massadata sisältää usein luottamuksellisia henkilötietoja. Niiden poistamisen tai anonymisoinnin vaikeus riippuu tietolajista. Anonymisoidustakin tiedosta voidaan tiedonlouhinnan menetelmin usein tunnistaa yksilöitä. Henkilötietoriski on este palvelujen ja liiketoiminnan kehittymiselle myös siksi, että kansalaisten luottamus tietojen pysymiseen salassa on (erityisesti muualla kuin Suomessa) varsin heikko¹³. **Riskin hallinta** vaatii lainsäädännön kehittämistä, kehittyneitä tietomalleja palvelujen tuottajille ja omadata-toimintamallin ja työkalujen käyttöönottoa.
- ii) **Tietoturvariskit.** Erityisesti suuret keskitetyt tietopankit houkuttelevat hyökkääjiä. **Riskin hallinta** vaatii tietoturvaosaamisen kehittämistä, hajautettujen tietomallien käyttöä, ja tiedonhallinnan kehittämistä niin, että julkaistava tieto on eri paikassa kuin salassa pidettävät tiedot.
- iii) **Turvallisuusriskit.** Erityisen tietoturvariskin muodostaa valtiolliseen tietoon kohdistuva urkinta. Suurimmilla ulkomaisilla tiedusteluorganisaatioilla on käytettävissään valtaiset resurssit järjestelmiin tunkeutumiseen. **Riskin hallinta** vaatii kotimaiseen tietoturvaosaamiseen panostamista.
- iv) **Teknologiariskit.** Massadatan, esineiden internetin ja älykkäiden kaupunkien teknologian kehitys on nopeaa, mutta tässä vaiheessa niille on tyypillistä kilpailevien, suljettujen ja yhteensopimattomien standardien rinnakkainen kehittäminen. Riskinä edelläkävijöille on väärin teknologioiden valinta ja jääminen teknologialoukkuun. **Riskin hallinta** vaatii palvelujen rakentamista avoimien datamallien, rajapintojen ja lähdekoodin varaan. Osariskinä tällöin on hajanainen toimintaympäristö ja keskitettyjen ratkaisujen kehittämisen vaikeus tai hitaus.
- v) **Toimittajariskit.** Nopeasti muuttuvassa ympäristössä on vaikea ennakoida toiminta- ja liiketoimintamallien lopullisia muotoja. Organisaatio voi jäädä vaille välttämätöntä tietoa tai toimittajaloukkuun, vaikka järjestelmä olisi avointa koodia tai tukisi avoimia rajapintoja. **Riskin hallinta** edellyttää palvelujen pidemmän aikavälin kehityskaaren ennakkointia ja skenaarioita, palvelualustoja, joista data on helposti siirrettävissä, ja sopimusmalleja, joiden ehtoja voidaan tarkistaa markkinan kehityksessä.

3.2 Massadatan tulevat kehityssuunnat

Massadatan hyödyntämiseen perustuvat palvelut ja liiketoiminta kehittyvät voimakkaasti.¹⁴ Big data käsitteenä syntyi noin kymmenen vuotta sitten. Ensimmäisen vaiheen hyödyntäjät olivat tyypillisesti suuria tietomääriä tuottavat Internet-yhtiöt (Google, Facebook, Yahoo, LinkedIn ym.) joiden kehitystyön pohjalta syntyivät monet käytössä olevat massadatateknologiat (mm. big data stack model, Hadoop, Apache Spark ym.). Tällöin rakennettiin massadataekosysteemin perusta.

Nyt käynnissä olevassa toisessa vaiheessa massadatamarkkinat ovat kypsymässä. Datan kokoaamisen perusinfrastruktuurista on siirrytty datan hyödyntämiseen ja analysointiin. Massadataa ovat alkaneet käyttää laajemmassa mittakaavassa myös muut kuin Internet-yritykset. Näiden haasteet massadatan hyödyntämisessä ovat monimutkaisemmat. Toisin kuin digitaalisena aikana syntyneet haastajat, ne eivät voi aloittaa nollasta. Olemassa olevia tuotantojärjestelmiä ei ole suunniteltu tuottamaan eikä hyödyntämään massadataa, eivätkä yritykset voi korvata liiketoiminnan kannalta keskeisiä järjestelmiä hetkessä. Massadatan hyödyntämiseen vaadittavien investointien kannattavuutta on vaikea arvioida.

¹³ David Meyer, Fortune, Europeans Remain Far from Sold on the Benefits of Big Data, 18.1.2016. <http://fortune.com/2016/01/18/europe-data/>

¹⁴ IDG Enterprise:n (2015) verkkokyselyyn vastanneista 1 139 yrityksestä yli neljäsosa oli toteuttanut dataperusteisia (data-driven) projekteja. Dataperusteisia projekteja toteuttaneiden yritysten määrässä oli kasvua 125 prosenttia edelliseen vuoteen verrattuna.

Muita hyödyntämisen esteitä ovat epäluottamus pilvipalveluita kohtaan, sekä varsin kirjava ja pk-yritysvetoinen palvelutarjoajien kenttä. Suuryrityksillä on usein vaikeuksia luottaa pieneen toimijaan kriittisen järjestelmän toimittajana. (Turck, 2016)

Seuraavaksi massadatan hyödyntäminen siirtyy sovelluksiin, sekä yritysten välisiin että yrityksiltä kuluttajille suunnattuihin palveluihin. Massadata alkaa samalla muuttua käsitteenä merkitykselliseksi sen käytön monipuolistuessa. Data on dataa, ja datan tehokas hyödyntäminen on kilpailuetu. Merkittäviä sovellusaloja ovat mm. e-kauppa ja markkinatutkimus, digitaalinen hallinto ja päätöksenteko, älykäs terveydenhuolto sekä turvallisuusbisnes. (Chen, Chiang ja Storey, 2013)

Teknologioiden kannalta tärkeitä käynnissä olevia massadatan kehityssuuntia ovat 1) avoimen koodin ja avoimien standardien kehittyminen, 2) tekoälyn ja massadatan yhdistäminen, sekä 3) yksityisten tietojen hallinta ja tietosuojaa massadataekosysteemeissä. Näistä tulevista kehityssuunnista tässä raportissa paneudutaan tarkemmin erityisesti kaupunkien massadataekosysteemin kehittämiseen sekä yksityisten tietojen hallintaan.

Kaupunkien massadataekosysteemin osalta esitetään nykyisiä avoimen datan ohjelmia laajempaa lähestymistapaa, kaupunkien datastrategiaa, mikä kattaa dataosaamisen lisäämisen, yksityisten data-lähteiden mukaan ottamisen, rajapintastrategian sekä yhteistyömallit datatoimijoiden kanssa kotimaassa ja ulkomailla.

Yksityisten tietojen osalta merkittävin ratkaistava kysymys on ihmisten mahdollisuus kontrolloida omaa dataansa. Sekä julkiset toimijat että yritykset (erityisesti eri alojen palvelualustat sekä ns. integraattoriyhtiöt, kuten Amazon, Apple, Google, Facebook ja Microsoft) kokoavat ja hallitsevat valtavaa määrää käyttäjien henkilökohtaista tietoa. Yritykset pyrkivät enenevässä määrin hyödyntämään hallussaan olevaa tietoa kaupallisesti. Niiden kokoamat henkilötiedot ja tietomassat ovat yrityksen omistamassa siilossa. Kriittiseksi kysymyksiä nousevat tiedon omistajuus, monikäyttöisyys ja hallinta. Euroopassa pyritään suuremmassa määrin kuin esim. Yhdysvalloissa vahvistamaan kansalaisten hallintaoikeuksia ja määräysvaltaa heitä koskevaan dataan.¹⁵

3.3 Massadatastrategia

Kyselyssä, haastatteluissa ja työpajoissa nousi selvästi esiin tarve suunnitella massadatan hyödyntämistä merkittävästi nykyistä systemaattisemmin. Yritykset tunnistivat massadatan mahdollisuudet, mutta suuri osa suhtautui niihin varsin passiivisesti. Kaupunkitoimijoiden suhtautuminen massadataan oli lähinnä ihmettelevä ja jäsentymätön.

Massadata ilmiönä tuo mahdollisuuksia lähes kaikille yhteiskunnan toimialoille liikenteestä finanssisektoriin ja terveydenhoitoon sekä kaikille organisaation toiminnan osa-alueille.

Massadata on terminä liian yleinen määritelläkseen toimintaa, ja alkaa pikkuhiljaa muuttua merkitykselliseksi (Chen, Chiang ja Storey, 2013). Yleistasoiset puheet ”tiedolla johtamisesta” ja ”parimmasta analyytiikasta” sisältävät nekin liian vähän konkretiaa tarjotakseen ainespuita datan aktiiviselle hyödyntämiselle. Organisaation on mahdotonta toteuttaa erittelemätöntä massadatastrategiaa yhtenä aktiivisena toimenpidekokonaisuutena.

Onnistuminen massadatan hyödyntämisessä vaatii strategian kohdentamisen, tavoitteiden kirkastamisen ja toimenpiteiden rajaamisen yhdeksi tai useammaksi hallittavan kokoiseksi kokonaisuudeksi. Massadatastrategian laatimisprosessin voi jakaa seuraaviin osa-alueisiin¹⁶: i) Massadatan tunnistaminen, ii) Tavoitteet massadatalle, iii) Massadatan käyttö: tiedonhankinta vs. tuotantokäyttö, ja iv) Massadatan käyttöönottonopeus.

¹⁵ European Commission: Protection of personal data. <http://ec.europa.eu/justice/data-protection/>

¹⁶ Massadatastrategiamalli: Davenport (2015, s. 59–84).

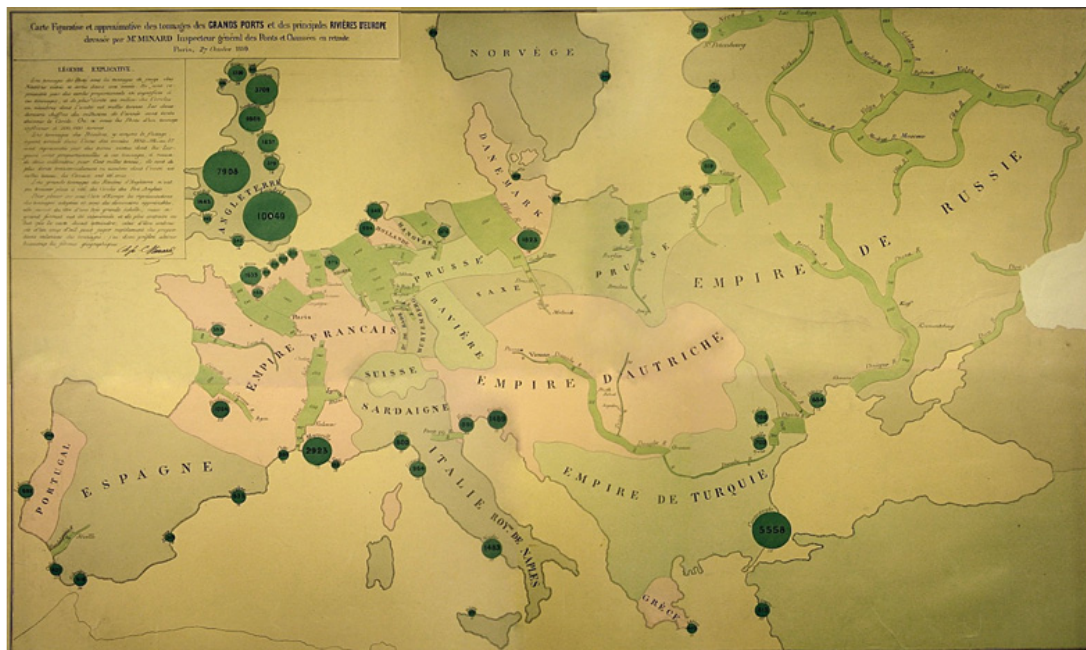
3.3.1 Massadatan tunnistaminen

Massadata on määritelmän mukaan dataa, jota on paljon, jota tulee nopeasti lisää ja joka on muodoltaan vaihtelevaa ja järjestymätöntä. Jotta sitä voidaan hyödyntää, pitää massadatastrategian laatiminen aloittaa organisaation massadatan tunnistamisesta ja priorisoinnista. Tässä voidaan käyttää apuna esimerkiksi seuraavaa taulukkoa, jonka avulla voidaan tehdä täsmällisempiä valintoja: ”massadatastrategian” sijaan organisaatio aikoo ”analysoida eri kanavia pitkin saatua kaupunkilaispalautetta saadakseen paremman kuvan asiakaspalveluprosessin toimivuudesta yli toimialojen” tai ”yhdistää ympäristö- ja liikennevirtadataa optimoidakseen jalankulkureittejä kaupungissa”.

Datan luonne	Datalähde	Toimiala	Toiminto
Suuri määrä dataa	Verkkopalvelut	Talousala	Markkinointi
Jäsentymätön data	Videolähteet	Turvallisuus	Hankintaketju
Datavirta	Sensorit	Ympäristöteknologia	Henkilöstö
Moniformaattidataa	Geenidata	Terveystieteet	Asiakkuuden hallinta
Linkitettyä dataa	Tilastotieto	Liikenne- ja kuljetusala	Maksuliikenne
Luottamuksellista dataa	Terveystiedot	Tuotteet	Tuotesuunnittelu
...

Datan hyödyntämisen kannalta on yhdentekevää, onko se ”isoa” vai ”pienää”. Läheskään kaikki toiminnan kehittämisen kannalta arvokkaat datalajit eivät ole määritelmällisesti massadataa, mutta ne voivat olla välttämättömiä osia tavoitteen saavuttamisessa. Nimikkeillä ei sinänsä ole väliä; massadata on vielä tällä hetkellä hype-termi, mutta se on vain uusi linkki jo vuosikymmenien tai -satojen ajan jatkuneeseen datan analysointiin pohjautuvien päätöksenteko- ja suunnittelumenetelmien ketjussa.

Kuvio 11 Euroopan suurimpien satamien ja jokien rahtimäärät (Charles Minard, 1850)



3.3.2 Tavoitteet massadatalle

Massadatan avulla voidaan tavoitella hyötyjä useammalla eri tavalla, kustannussäästöistä tuotekehitykseen. Hyödyt jakautuvat **kustannussäästöihin, prosessien nopeutumiseen, uusiin tuotteisiin tai palveluihin sekä parempaan päätöksentekoon.**

Kustannussäästöt ovat yleisin massadatan hyödyntämisen tavoite suurilla datamääriä käyttäville organisaatioille ja toiminnan osa-alueille (esim. talous, ympäristö). Massadatakusterit ovat perinteisiä tietokantoja oleellisesti halvempi tapa hallita suuria tietomääriä. Kustannusero massadataratkaisun ja perinteisen relaatiotietokannan välillä voi olla jopa 20–30-kertainen. Päätökset perustuvat suhteellisen suoraviivaisiin panos–tuotos-laskelmiin, eivätkä suoranaisesti vaikuta organisaation perustoimintaan. Suurimpia huomioon otettavia asioita ovat palvelun saatavuus ja käytettävyyys.

Prosessien nopeuttaminen on toinen yleinen massadatan hyödyntämisen tavoite, erityisesti monimutkaisissa analyysiprosesseissa. Esimerkiksi monilla palvelualoilla (mm. matkailupalvelut) hinnat muuttuvat nopeasti markkinatilanteen mukaan. Nopeammat prosessit mahdollistavat myös suuremman määrän vaihtoehtoja; analytiikasta voidaan ajaa jopa satoja eri variantteja 10–20 sijaan. Ajansäästöillä on kuitenkin oltava selkeät tavoitteet, jotka on syytä suunnitella ennakkoon. Keskeistä on määritellä tavoitellaanko...

- parempaa prosessien ja tehokkuuden ymmärrystä suuremmalla määrällä eri malleja,
- useammin tehtävää toiminnan suunnan ja parametrien tarkistusta,
- suurempaa määrää dataa ja muuttujia tarkemman kuvan saamiseksi vai
- nopeampaa reagointia toimintaympäristön muutoksiin.

Uudet tuotteet ja palvelut ovat kenties kunnianhimoisin massadatan sovellus. Monien verkko-palveluyritysten uudet innovaatiot perustuvat pitkälle näihin (Google, sosiaalisen median alustat kuten LinkedIn, Facebook, mediapalveluiden (Spotify, Netflix ym.) suositusalgoritmit). Kaupunkien ja muiden julkisen sektorin toimijoiden osalta uusia palveluita voivat olla esimerkiksi käyttäjien mukaan räätälöidyt palvelut. Suomen veroehdotus on tästä varhainen esimerkki. Oppivat järjestelmät yhdistävät massadata-alustoja, joukkoistettua tiedon tuottamista ja tekoälyä. Tuore esimerkki ovat Teslan oppivat autot¹⁷, jotka täydentävät itseajavien autojen tekoälyä ja karttatietoja ihmiskuljettajien ajotiedoilla.

Tuotekehityspotentiaalin hyödyntäminen vaatii sille nimetyn massadataympäristön ja -toimintamallin perustamista. Kaupunkiympäristöissä esimerkiksi paikallisesti rajatut palveluympäristöt voivat mahdollistaa uusien tuotteiden ja palveluiden kehittämisen ja testaamisen todellisissa tuotanto-olosuhteissa. Tämä on lähtökohtana monissa älykaupunkitoteutuksissa, kuten Helsingin Kalasatamassa.¹⁸

Parempi päätöksenteko on ollut datan hyödyntämisen ja analytiikan tavoite teollistumisen alkua ajoista lähtien. Koska massadatan käsittelyteknologia on mahdollistanut datan kokoamisen ja analytiikan paljon aikaisempaa suuremmista ja järjestäytymättömmistä kohteista, on tietopohjaisen päätöksenteon piiriin tullut paljon toimialoja, joihin analytiikkaa ei aiemmin voitu vastaavassa määrin soveltaa. Esim. infrastruktuureista (rakennukset, rakenteet, kadut, kaupungin verkostot) voidaan koota massadataa sensorien avulla ja seurata niiden tilaa reaaliaikaisesti. Asiakaspalveluprosesseja voidaan seurata verkkopalvelujen klikkausten ja puheen automaattisen analysoinnin avulla.

Kaupungeissa massadatan ja joukkoistetun tiedonhankinnan yhdistäminen mahdollistaa esimerkiksi aiempaa olennaisesti täsmällisemmät palautejärjestelmät sekä ympäristön turvallisuutta parantavan ratkaisut.

3.3.3 Massadatan käyttö: tiedonhankinta ja tuotantokäyttö

Massadatan käytön tärkeimmät vaiheet tai toimintamoodit ovat tiedonhankinta (massadatasta oppiminen ja ymmärryksen lisääminen) ja tuotanto (massadatan käyttö osana prosessien, tuotteiden ja palveluiden toteutusta). Tiedonhankinnassa tarvitaan erilaista osaamista, rakenteita, välineitä ja prosesseja kuin datan tuotantokäytössä.

Tiedonhankinta on ollut analytiikan tavoitteena aina, mutta massadatan mukanaan tuomat johtamishaasteet ja liiketoimintamahdollisuudet ovat kasvattaneet sen merkitystä. Tiedonhankinta toteutetaan organisaation innovaatio-, tuote- ja palvelukehitys- tai tutkimusyksiköissä, ei niinkään IT-osas-

¹⁷ <http://fortune.com/2015/10/16/how-tesla-autopilot-learns/>

¹⁸ <http://fiksukalasatama.fi/>

tolla. Tiedonhankinta toimii, kun sitä tukee rohkea kokeilukulttuuri ja lupa epäonnistua. Tämä voi olla haastavaa julkisen sektorin organisaatioille, ja erityisen haastavaa esimerkiksi terveydenhuollon alueella. Datalaboratoriot ja sandboxit auttavat tarjoamalla kokeiluille turvallisen ympäristön.

Tiedonhankinnan tuloksena syntyy ideoita, tuote- ja palveluominaisuuksia tai parannusehdotuksia olemassa oleviin tuotteisiin, palveluihin ja prosesseihin. Pienetkin muutokset ovat tärkeitä prosesseja optimoitaessa.

Tuotantokäytössä data integroidaan tuotteisiin ja prosesseihin. Skaalautuvuus-, luotettavuus- ja turvallisuuskysymykset on ratkaistava. Yleensä vain osa tiedonhankinnassa löytyneistä mahdollisuuksista sopii tuotantokäyttöön johtuen mm. strategisista linjauksista, organisaatiokulttuurista tai prosessien erilaisuudesta tai huonosta panos-tuotossuhteesta. Jos mitkään ideat eivät tule käyttöön, organisaation tuotantomekanismit ovat liian jäykät tai teknisesti rajoittuneet. Tuotantokäytössä datan sovellusosaajat ovat profiililtaan erilaisia kuin tiedonhankinnassa. Heidän osaamisensa on erityisesti integraatiossa, soveltamisessa ja optimoinnissa. He ovat tyypillisemmin IT-osaajia kuin liiketoiminnan kehittäjiä.

3.3.4 Massadatan käyttöönottonopeus

Mikä on oikea hetki uusien asioiden käyttöönottoon? Toimintaympäristöstä ja organisaatiosta riippuen vastaus vaihtelee. Joskus on liikuttava aggressiivisesti, joskus kannattaa odottaa. Haasteena on luotettavuuden ja innovatiivisuuden välisen tasapainon löytäminen. Seuraavia Davenportin (2014) esittämiä mittareita voi käyttää apuna suunnittelussa:

Organisaation on syytä olla varovainen, jos

- muut toimijat alalla eivät juurikaan hyödynnä massadataa
- teknologia ei ole ollut aiemmin toimialalla ratkaisevassa roolissa
- käytettävissä ei ole tarpeeksi dataa asiakkaista tai muista kriittisistä tuotantotekijöistä
- organisaatiolla ei ole muutuskulttuuria, vaan se on yleensä seurannut trendejä hieman jäljessä.

Organisaation on hyvä olla ripeä, jos

- Toimiala hyödyntää jo massadataa tai analytiikkaa
- Organisaatio haluaa pysyä kilpailun edellä
- Organisaatiolla on teknologiaa ja dataa hyödyntävä kulttuuri
- Talosta löytyy osaamista massadatan käyttöön.

On liikuttava heti ja nopeasti, jos

- Kilpailijat hyödyntävät massadataa aggressiivisesti
- Organisaatiolla on aiempaa menestystä analytiikan hyödyntämisestä
- Organisaatio on tottunut käyttämään teknologiaa toimialansa uudistamisessa
- Talosta löytyy kaikki tarvittava osaaminen.

3.3.5 Massadata-valmiusindeksi

DELTA-malli (Data, Enterprise, Leadership, Targets, Analysts) on varsin yleisesti käytetty työkalu analytiikkavalmiuksien ja onnistumismahdollisuuksien arviointiin. Työkalu on päivitetty massadataan sopivaksi DELTA-malliksi lisäämällä mukaan toinen T, Technology.

Organisaatio voi toteuttaa Massadata-valmiusarvioinnin seuraavan kriteeristön avulla¹⁹;

¹⁹ Kriteeristö: Davenport (2015, s. 206–209).

DATA

- Käytettävissämme on tarpeeksi suurta, jäsentymätöntä, nopeasti muuttuvaa ja toiminnalle relevanttia dataa
- Integroimme dataa eri lähteistä keskitettyyn datapankkiin
- Yhdistelemme sisäistä ja ulkoista dataa mahdollistaaksemme toiminnan syvällisen analyysin
- Käytämme metadataa ja standardeja datassa, jota käytämme analysointiin
- Käyttäjät, päättäjät ja tuotekehittäjät luottavat datamme laatuun

ORGANISAATIO

- Yhdistelemme perinteistä analytiikkaa ja massadata-toimintamalleja
- Johto varmistaa, että eri yksiköt tekevät yhteistyötä massadatan ja analytiikan toimintamallien kehittämiseksi ja priorisoimiseksi
- Käytämme analyytikkoja ja datatieteilijöitä poikkialaisesti koko organisaatiossa
- Analytiikka- ja massadatatoiminnot ovat riittävästi resursoituja
- Teemme yhteistyötä yli yksikköjen sekä asiakkaiden ja kumppaneiden kanssa jakaaksemme massadataa

JOHTAMINEN

- Ylin johto on perillä massadatasta ja arvioi aktiivisesti sen mahdollisuuksia
- Ylin johto sparraa eri yksiköjä ja keskijohtoa massadatan ja analytiikan käytössä
- Ylin johto käyttää massadataa ja analytiikkaa johtamisen ja strategian välineenä
- Keskijohto käyttää massadataa ja analytiikkaa päätöksenteon apuna
- Ylin johto seuraa ja arvioi massadataprosesseja ja -resursseja (ihmiset, teknologia, data)

KOhteet

- Kehitämme massadatasta korkean lisäarvon mahdollisuuksia erottuaksemme kilpailijoista
- Massadataa hyödyntävät tuotteet ja palvelut ovat olennainen osa innovaatioprosessiamme
- Tunnistamme sisäiset massadatan käyttömahdollisuudet evaluoimalla prosessejamme, strategiaamme ja toimintaamme
- Teemme säännöllisesti datavetoisia kokeiluja
- Arvioimme päätöksiemme onnistuneisuutta uuden massadatan ja analytiikkatiedon valossa

TEKNOLOGIA

- Olemme tutkineet tai käyttäneet rinnakkaisen tiedonkäsittelyn välineitä (mm. Hadoop) massadataan
- Käytämme datavisualisointia johtamisen ja päätöksenteon välineenä
- Olemme tutkineet tai käyttäneet pilvipalveluita datan prosessointiin
- Olemme tutkineet ja käyttäneet avoimen lähdekoodin ratkaisuja dataan
- Olemme tutkineet tai käyttäneet välineitä prosessoida jäsentymätöntä dataa, kuten tekstiä tai kuvia

ANALYTIikka

- Meillä on datatieteilijöitä ja analytiikkaosaajia tarpeeksi
- Ylin johto luottaa datatieteilijöihimme ja analytiikkaosaajiimme
- Datatieteilijät ja analytiikkaosaajat ymmärtävät prosessejamme
- Datatieteilijät, analytiikkaosaajat ja IT-johto työskentelevät tehokkaasti tiimeissä
- Meillä on kehitysohjelmat datatieteilijöiden ja analytiikkaosaajien taitojen kehittämiseksi

3.4 Omadata- ja blockchain-teknologiat

Omadata²⁰ on erityisen kriittisessä asemassa massadata -maisemassa. Yksityistä dataa voi sisältyä lähes mihin tahansa massadata -aineistoon, ja sen puutteellinen hallinta toisaalta synnyttää merkittäviä tietoturvariskejä, toisaalta voi estää erittäin käyttökelpoisen datan käytön.

Omadatan hyödyntäminen voi olla monimutkaista; datan käyttö muuttuu samaa tahtia nopeasti kehittyvän teknologian kanssa. Lakien ja asetusten tasolla on mahdotonta varautua ennakkoiden kaikkiin niihin eri tapoihin, millä yksityistä dataa palveluissa tuotetaan, käsitellään ja jaetaan. Tärkein osa omadata-ajattelua onkin painotuksen siirtäminen viranomaisen toteuttamasta datan hallinnasta ja **lainsäädännöstä** käyttäjien itsensä hallitsemiin datan hallinnan **työkaluihin**. Omadatan pääajatus on tarjota käyttäjälle itselleen konekielinen pääsy dataansa ja välineet datan käytön lupien ja suostumusten hallintaan.

Tässä tilanteessa vahva panostus omadataan on äärimmäisen tärkeää. Euroopassa tarvitaan sekä lainsäädännöllisiä että teknisiä keinoja tämän mahdollistamiseksi. Onnistuminen omadatasuon on myös taloudellinen kysymys. Tietoturvaohjelmien ja -vuotojen takia erityisesti Euroopan maiden kansalaiset ovat enenevässä määrin epäluuloisia yritysten ja hallitusten kykyyn suojata luottamuksellista tietoa²¹. Ilmiö voi hidastaa verkkopohjaisten palvelujen kasvua ja massadatasta toivottujen hyötyjen saamista.

Kenties kiinnostavimpia teknologioita turvallisten, hajautettujen omadata- ja autentikointiratkaisujen toteuttamiseen ovat lohkoketju- eli blockchain-ratkaisut, joita on käytetty eniten virtuaalivaluutoissa (bitcoin ym.). Lohkoketjut sopivat erittäin monipuolisesti mm. erilaiseen autentikointiin, tiedon hallittuun käsittelyyn, turvaratkaisuihin ja ylipäättään jakamiseen, on se sitten laskentatehoa, tallennustilaa tai kaistaa. Viro on suojaamassa blockchainilla kaikkien kansalaisten terveystietoja²².

Mattila ja Seppälä (2015) esittävät, että blockchain-teknologioiden muutosvoima on samaa luokkaa kuin internetin taustalla olevien www/TCP-IP-teknologioiden. Massadata-kontekstissa mahdollisuudet ovat merkittävät. Yksi massadatan hyödyntämisen este ovat järjestelmien kompleksisuuden mukana nousevat kustannukset; suuren datamäärän hallinta keskitetysti kuluttaa valtavasti resursseja. Blockchain-teknologioilla voidaan toteuttaa samankaltaisia operaatioita hajauttaen tarvittavat prosessit kaikille järjestelmässä mukana oleville laitteille. Tästä syystä lohkoketjuteknologiat ovat kehittyessään kustannustehokas tapa toteuttaa monimutkaisia järjestelmiä.

Blockchainin perusteet on kuvattu hyvin ETLAn viime vuonna julkaisemassa Matilan ja Seppälän (2015) raportissa:

Blockchain- eli lohkoketjuteknologialla tarkoitetaan menetelmää, jonka avulla toisilleen tuntemattomat tahot voivat tuottaa ja ylläpitää lähestulkoon mitä tahansa tietokantoja täysin hajautetusti keskenään. Menetelmä toimii siten, että jokaiselle halukkaalle taholle jaetaan kopio tietokannasta (tai sen osasta) ja kaikki osalliset voivat myös tehdä tietokantaan muutoksia yleisesti hyväksytyjen sääntöjen puitteissa. Eri tahojen kulloinkin tekemät muutokset tallennetaan tietokantaan kootusti tietyin väliajoin yhteen niputettuina paketteina, joita kutsutaan lohkoiksi. Kun uusia lohkoja liitetään alkuperäiseen tietokantaan, muodostuu niistä lohkoketju, eli ajantasainen tietokanta kaikkine muutoksineen.

Lohkoketjuteknologia vaikuttaa lupaavalta vaihtoehdolta älykkäiden tuotteiden ja palveluiden platformiteknologiaksi tilanteissa, joissa muut arkkitehtuuriratkaisut olisivat liian kustannusintensiivisiä tuotannon kokonaiskustannuksiin suhteutettuna. Lohkoketjuteknologian avulla älykkäiden tuotteiden ja palveluiden verkosto voidaan tuottaa täysin itsenäisenä vertaisverkkona laitteiden itsensä välillä, eikä erillistä ulkoisiin pilvipalveluihin perustuvaa tausta-arkkitehtuuria tarvita periaatteessa lainkaan.

20 Omadatasuon on tehty vuonna 2014 opas (Poikola, Kuikkaniemi ja Kuittinen, 2014).

21 David Meyer, Fortune, Europeans Remain Far from Sold on the Benefits of Big Data, 18.1.2016. <http://fortune.com/2016/01/18/europe-data/>

22 Daniel Palmer, Coindesk; Blockchain Startup to Secure 1 Million e-Health Records in Estonia. <http://www.coindesk.com/blockchain-startup-aims-to-secure-1-million-estonian-health-records/>

Case: D-CENT

D-CENT^a (Decentralised Citizens ENGagement Technologies) on eurooppalainen avointen kansalaisosallistumisen välineiden kehittämishanke. Se tarjoaa digitaalisia työkaluja demokraattiseen vaikuttamiseen ja osallistumiseen hallinnolle, kansanliikkeille ja kansalaisjärjestöille. Hajautettua, avointa alustaa kehitetään avoimen lähdekoodin ratkaisuna yhdessä käyttäjien kanssa. Kokemuksia kehitystyöhön kerätään Suomesta, Espanjasta ja Islannista, joissa kaikissa on jo käytössä suoran verkkodemokratian työkaluja (esim. Avoin ministeriö).

D-CENT pyrkii kehittämään tietoturvallisen ja kestävä vaihtoehdon kaupallisille alustoille. D-CENT haluaa haastaa Internetin nykyisen kehityksen, jossa valta ja markkinaosuudet ovat keskittyneet muutamille suurille toimijoille.

D-CENT:n Suomen pilotissa on kehitetty palvelua, jolla kansalaiset voivat seurata Helsingin kaupungin päätöksentekoa valitsemillaan hakusanoilla sekä yhteiskirjoittaa porukalla dokumentteja – esimerkiksi tietopyynnön virkamiehelle tai tiedotteen medialle. Tieto ja dokumentit haetaan Open Ahjo -rajapinnasta.

Hankkeessa on tehty runsaasti avoimesti jaettua tutkimusta, laajaa standardointityötä (W3C, IETF), tuotekehitystä avoimella lähdekoodilla (GitHub), blockchain-ratkaisuihin perustuvia autentikointityökaluja sekä avointa FreeCoin -bittivaluutta.^b

2,5 vuotta kestävä D-CENT-hanke päättyy toukokuun lopussa 2016. Mukana on yhdeksän organisaatiota Espanjasta, Hollannista, Irlannista, Islannista, Isosta-Britanniasta, Ranskasta ja Suomesta. Forum Virium Helsinki ja Avoin ministeriö ovat mukana Suomesta..

^a D CENT: www.dcentproject.eu

^b D CENT tools: <http://tools.dcentproject.eu>

Blockchain on vielä ilmiönä melko tuore ja teknologiat vasta kehittymässä. Nyt on oikea hetki satsata niiden tutkimus- ja kehitystoimintaan. Tärkeimpiä sovellusalueita ovat autentikointi ja yksityisyyden suoja. Blockchain-menetelmillä voidaan toteuttaa omadatan tarvitsema hajautettu mutta varma tunnistaminen. Blockchainin mahdollisuuksia kannattaa myös tutkia turvallisten eDemokratia-ympäristöjen, hajautettujen päätöksentekomekanismien ja hajautettujen käyttäjien välisen vaihdannan ja arvonaluonnin ympäristöjen toteuttamisessa.

Suomessa on jo käynnissä kokeiluja, joissa ko. teknologioita hyödynnetään ja testataan. Tulokset ovat lupaavia. Hankkeista kenties kiinnostavin on kansainvälinen D-CENT, minkä ratkaisuja on pilotoitu Helsingissä.

3.5 Päätelmät massadatan hyödyntämisestä

Avoin data, massadata ja omadata sekä niitä hyödyntävät sovellusalueet kuuluvat kaikki samaan, digitaalisuuden ja datan varaan rakentuvaan älypalvelujen ja älykkään infrastruktuurin kokonaisuuteen. Erilliset strategiat ja toimenpideohjelmat digitaalisuuden eri osa-alueille ovat usein tarpeen, jotta asioita voidaan edistää rahoitusmahdollisuuksien reunaehtojen puitteissa. Näiden ohella maassamme olisi oltava myös kansallinen, erillisohjelmien yli ulottuva datastrategia, missä rinnakkaisia ilmiöitä tarkastellaan kokonaisuutena suhteessa eurooppalaiseen ja globaaliin kehitykseen sekä suunnitellaan ohjelma-, toimija- ja toimialarajat ylittävää dataosaamisen, datan saatavuuden ja datan käytön edistämistä.

Luvun kirjoittajan keskeiset näkemykset ekosysteemin kehittämisen kannalta ovat:

- Massadatamarkkinat ovat kypsymässä ja siirtymässä perusinfrastruktuurista datan hyödyntämiseen ja analysointiin. Perinteisten yritysten haasteet massadatan hyödyntämisessä ovat kuitenkin monimutkaisemmat kuin alan pioneereina toimineiden Internet-yritysten.
- Massadatan hyödyntäminen on siirtymässä sovelluksiin. Massadatan hyödyntämiseen vaadittavien investointien kannattavuutta on vielä vaikea arvioida.

- Massadata on laaja käsite. Onnistuminen massadatan hyödyntämisessä vaatii strategian kohdentamisen, tavoitteiden kirkastamisen ja toimenpiteiden rajaamisen yhdeksi tai useammaksi hallittavan kokoiseksi kokonaisuudeksi.
- Hyödyt jakautuvat kustannussäästöihin, prosessien nopeutumiseen, uusiin tuotteisiin tai palveluihin sekä parempaan päätöksentekoon.
- Massadatan käytön tärkeimmät toimintamoodit ovat tiedonhankinta (massadatasta oppiminen ja ymmärryksen lisääminen) ja tuotanto (massadatan käyttö osana prosessien, tuotteiden ja palveluiden toteutusta). Näissä tarvitaan erilaista osaamista, rakenteita, välineitä ja prosesseja.
- Omadata on kriittisessä asemassa massadata-maisemassa. Omadataa voi sisältyä lähes mihin tahansa massadata-aineistoon. Sen huono hallinta voi hidastaa verkkopohjaisten palvelujen kasvua ja massadatasta toivottujen hyötyjen saamista.
- Haasteiden ratkaisemiseen tarvitaan uusia toimintamalleja, kuten omadata-ekosysteemiä, missä painotus siirtyy lainsäädännöstä käyttäjien hallitsemiin datan hallinnan työkaluihin.
- Yksi lupaavimmista mahdollisuuksista omadata-työkalujen kehittämiseen ovat blockchain/lohkaketju-teknologiat, jotka sopivat hajautettuun autentikointiin, tiedon käsittelyyn, turvartkaisuihin sekä arvon tai resurssien jakamiseen.

4 Massadatan hyödyntäminen julkisissa palveluissa

Massadatan hyödyntämiseen liittyvien julkisten palveluiden kehittämispotentiaali ja kehittäjien kiinnostus on suuri. Massadatan avulla voitaisiin nopeuttaa prosesseja, luoda uusia tuotteita tai palveluita ja tehdä parempia päätöksiä. Riskien (tiedon väärinkäyttö, omistajuus, avoimuuden kysymykset) vaikea ennustettavuus kuitenkin estää datan laajamittaista käyttöä. Tämä luku esittelee massadatan käyttöä julkishallinnossa ja kaupungeissa sekä eri toimijoiden roolia. Luku valottaa esimerkinomaisesti liikkumis- ja hyvinvointipalveluiden kehittämiseen liittyviä ajankohtaisia kysymyksiä ja pullonkauloja Suomessa ja tuo esille kansainvälisiä käytäntöjä.

4.1 Massadatan lähteet ja tavoitteet massadatan käytölle julkishallinnossa

Julkisia palveluita tuottavat mm. valtio, kunnat, alueet ja kuntayhtymät, julkisomisteiset yritykset ja palvelulaitokset (Kela ym.) sekä yksityiset tahot palvelutuottajasopimusten kautta. Kaikki julkiset tahot voivat käyttää massadataa palvelujen parantamisen, tiedolla johtamisen ja kilpailukyvyn parantamisen välineenä. Hyödyt ovat samat sekä yrityksille että julkisille tahoille: **kustannussäästöt, prosessien nopeutuminen, uudet tuotteet tai palvelut sekä parempi päätöksenteko.**

Julkisen sektorin tuottamaa massadataa syntyy monesta eri lähteestä:

- Tilastoista: aikasarjat mm. väestötiedoista, ympäristötiedoista, rakentamisen tiedoista, av-aineistoista (kuvat ja videot). Tilastodata voi sisältää luottamuksellisia henkilötietoja.
- Palveluprosesseissa: mm. asiointitiedot, koulutiedot, terveystiedot, sosiaalipalvelujen tiedot. Data sisältää usein luottamuksellisia tietoja.
- Päätöksenteossa: mm. suunnitteluasiakirjat, tausta-asiakirjat, luvat, kuva-aineisto, hallinnon päätöksiä. Data saattaa sisältää luottamuksellisia tietoja.
- Infrastruktuurista: mm. liikenne-, vesi-, energia-, puhtaanapito-, turvallisuus- ym. järjestelmät, sensoriverkot, paikkatieto. Data ei yleensä sisällä luottamuksellista tietoa.

Julkishallinnon palveluiden kehittäminen ja tuottaminen perustuvat entistä enemmän datan hyödyntämiseen. Kuten kaupallisessa palvelutuotannossa myös julkishallinnossa tieto on alusta, jonka päälle rakennetaan asiakaslähtöisiä ja kustannustehokkaita palveluita, jotka perustuvat käyttäjien ja sidosryhmien tarpeisiin sekä odotuksiin. Asiakaslähtöisyyden ydinedellytys on, että ymmärretään kansalaisen tarpeet ja kohdennetaan yhteiskunnan resurssit julkissektorin tietokehittämisessä oikeisiin asioihin.

Julkisten palveluiden massadata tuottaa hyötyä yksittäisille palvelutuottajille palvelujen tuottavuuden, tehokkuuden, seurattavuuden ja mitattavuuden kautta, ja palvelutuottanto-organisaatioille ja -verkostoille prosessien, käytäntöjen ja tiedon tehostumisen, läpinäkyvyyden ja jaettavuuden kautta. Palvelujen käyttäjät voivat hyötyä massadatan käytöstä palvelujen laadun, saatavuuden, räätälöinnin ja hallittavuuden paranemisen kautta. Yrityskäyttäjät hyötyvät palveluprosessien avaamisen ja monipuolistumisen kautta. Tällöin erityisesti korkean arvonlisän yritysten mukana olo datan avaamisessa, jalostamisessa ja hyödyntämisessä palveluiden rakentamiseen on tärkeää. Tutkijoille julkisten palveluiden massadata tuo hyötyä datan saatavuuden, laadun ja käytettävyyden paranemisen kautta. Tutki-

mustiedon paraneminen tuottaa puolestaan parempaa tietoa palvelujen laadun parantamiseen, jolloin syntyy positiivinen kierre.

Datalähtöisempää julkishallintoa voidaan yleisesti tarkastella esimerkiksi kolmella seuraavalla osa-alueella: datalähtöisen päätöksenteon ja jatkuvan organisaatiokehityksen tuominen osaksi julkishallinnon kulttuuria, suorituskyvyn parantaminen ja kansalaisten henkilökohtaisemmat palvelut. Julkishallinnon toimijoiden datalähtöisen päätöksenteon tehostamisen ja suorituskyvyn optimoimisen lisäksi massadatan avulla voidaan tehostaa samankaltaisten organisaatioiden ja yksiköiden keskinäistä suorituskäytännöiden vertailua (organizational benchmarking).

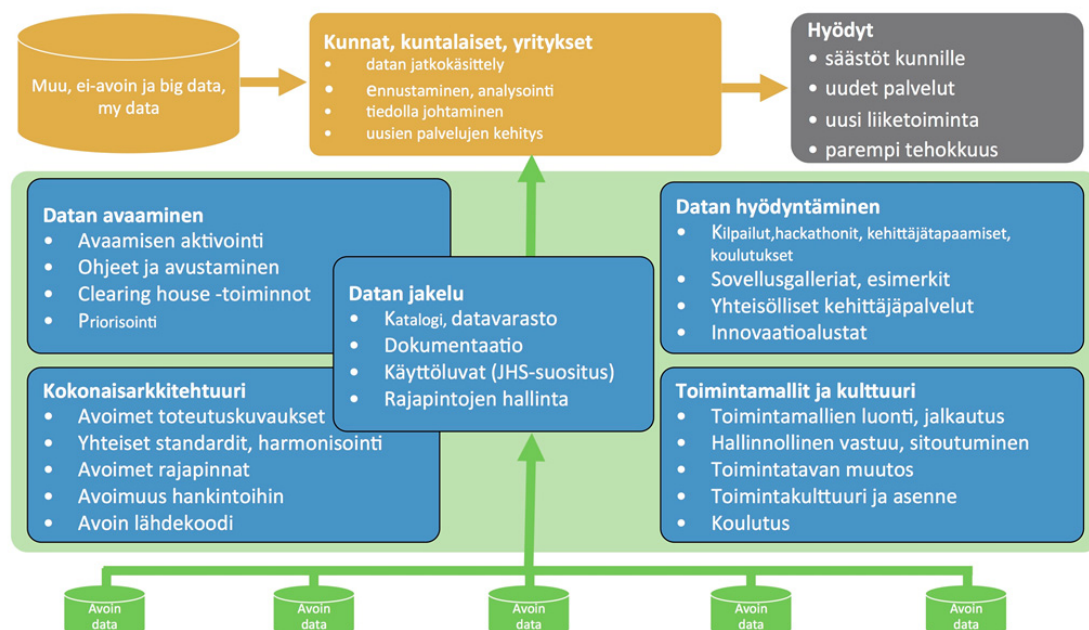
4.2 Kaupungit ja massadatan hyödyntäminen

4.2.1 Massadata ja 6Aika -strategia

Raportin julkisia palveluita koskeva osio toteutettiin yhteistyössä Suomen kuuden suurimman kaupungin (ns. kuutoskaupungit; Helsinki, Espoo, Vantaa, Tampere, Turku ja Oulu) yhteisen 6Aika-strategian Avoin data ja rajapinnat -kärkihankkeen kanssa. Yksikään kaupunki tai edes kuutoskaupungit yhdessä eivät kuitenkaan muodosta kriittistä massaa globaaleilla digitaalisten palvelujen markkinoilla. 6Aika-ohjelman kokemusten valossa jo keskisuurten kaupunkien osalta yksittäinen kaupunki on aivan liian pieni toimiakseen kunnolla uusien innovaatioiden alustana, pienemmistä kunnista puhumattakaan. Tavoitteena onkin pyrkiä rakentamaan Suomesta yhtenäinen massadatan koelaboratorio. Case-esimerkit on valittu kaupungeista, mutta tavoitteena on kaikkien julkisten toimijoiden ja yksityisten yritysten välisen yhteistyön maksimoiminen. Suositukset soveltunevat moniin kuntiin ja julkishallinnon organisaatioihin sekä niiden kanssa yhteistyötä tekeviin yrityksiin ja muihin toimijoihin.

6Aika on kestävä kaupunkikehityksen strategia, jota toteuttavat yhteistyössä Suomen suurimmat kaupungit: Helsinki, Espoo, Vantaa, Tampere, Turku ja Oulu. Vuosina 2014–2020 toteutettavan strategian tavoitteena on synnyttää Suomeen uutta osaamista, liiketoimintaa ja työpaikkoja. 6Aika on kansainvälisestikin ainutlaatuinen strategia, joka voi toimia merkittävän massadataosaamisen ja -liiketoiminnan vauhdittajana Suomessa. Strategian toisen, vuonna 2017 alkavan vaiheen valmistelu on nyt käynnissä. Sen suunnittelussa tulisi tehdä mahdollisimman laajaa yhteistyötä sekä kansallisten ohjelmien (Massadata, omadata, MaaS ym.) että kansainvälisten strategiavalmistelujen (Digital Single Market ym.) kanssa.

Kuvio 12 Massadatan hyödyntäminen kunnissa



Strategiassa on kolme painopistealuetta. Avoimet innovaatioalustat -painopistealueella luodaan ja testataan uusia palveluja ja tuotteita aidossa olosuhteissa. Avoin data ja rajapinnat -painopistealueella avattu data toimii uusien palvelujen sekä toiminnan kehittämisen raaka-aineena. Avoin osallisuus ja asiakkuus -painopistealue kokoaa kaupunkiyhteisön kehittämään yhdessä palveluinnovaatioita ja asiakasprosesseja.

Painopistealueilla toteutetaan kolmivuotiset, kaikkien kaupunkien yhteiset kärkihankkeet, joiden avulla rakennetaan perustaa innovaatiotoiminnalle ja luodaan edellytykset yhteistyölle. Kärkihankkeita tuetaan pilotti- ja kokeiluhankkeilla, jotka käytännössä testaavat ja kehittävät edelleen kärkihankkeiden sisältöjä. 6Aika-strategian hankkeet ovat aina vähintään kahden toimijan yhteishankkeita. Näitä toimijoita voivat kaupunkiorganisaatioiden lisäksi olla muut julkiset tai kolmannen sektorin toimijat.

4.2.2 Massadatan suuret mahdollisuudet

Massadatan hyödyntämisen nykykäytännöistä suomalaisissa kaupungeissa²³ ja yrityksissä kerättiin aineistoa ja näkemyksiä haastatteluilla ja kolmella työpajalla. Haastatteluja tehtiin yhteensä 22 kappaletta ja ne toteutettiin puhelimitse marraskuussa 2015. Yrityksiä ja julkisia tahoja haastateltiin massadatan hyödyntämisen nykykäytännöistä ja toimivista esimerkeistä. Näiden lisäksi haastateltavilta kerättiin näkemyksiä massadatan käytön oikeudellisista kysymyksistä (mm. omistajuus, hallinta, vastuu). Haastatteluaineistoa syvennettiin kolmessa työpajassa marras-joulukuussa 2015 Oulussa, Tampereella ja Helsingissä. Työpajoihin osallistui kaupunkien ja yritysten edustajia. Niissä pohdittiin massadatan mahdollisuuksia ja haasteita erityisesti omadatan näkökulmasta. Pajojen erityisinä kohdealueina olivat liikkuminen palveluna (MaaS, Mobility as a Service), sosiaali- ja terveyspalvelut sekä kauppa.

Massadatan käytöllä on suuret mahdollisuudet yhteiskunnan tasolla mm. viestinnässä, liikkumisessa, kuljetuksessa, energiasektorilla sekä ympäristöpalveluissa. Yksilötasolla mahdollisuudet liittyvät mm. oman käyttäytymisen tai kuluttamisen seurantaan. Tätä tutkimusta varten tehdyissä haastatteluissa keskityttiin erityisesti liikkumisen ja sosiaali- ja terveyspalveluiden massadataan suomalaisissa kaupungeissa. Etlan tekemän kyselyn²⁴ mukaan yritykset hyödyntävät massadataa eniten päätöksenteossa, markkinoinnissa ja myynnissä sekä tavaroiden ja palveluiden tuottamisessa. Kaupunkien näkemykset massadatan hyödyntämisestä ovat samansuuntaisia: massadataa hyödynnetään eniten palveluiden tuottamisessa ja päätöksenteossa. Näiden lisäksi tulevaisuuden mahdollisuuksina pidetään datan keruupaikkojen ja -välineiden laajentumista ja reaaliaikaistumista, potentiaalia tuottaa ja jalostaa tietoa vielä laadukkaammin sekä yksityishenkilöiden että julkisten yhteisöjen, kuten kaupungin tai valtion päätöksenteon tueksi, datan määrän kasvun kautta mahdollistavaa analytiikkaa sekä datalajien yhdistämisen ja kokemusdatan kautta syntyvää uutta tietokerrosta.

4.2.3 Selkeät roolit, kapeat pullonkaulat

Haastattelujen perusteella on ilmeistä, että kaupungit, yritykset ja esimerkiksi tutkimuslaitokset näkevät itsellään erilaiset roolit massadatan keruussa, jalostamisessa ja käytössä. Roolit näyttävät muodostuvan nykyisen massadatan käytön rajojen perusteella: kaupungit näkevät pääasialliseksi tehtäväkseen massadatan käytön mahdollistamisen ja datan jakamisen avointen rajapintojen kautta, yritykset näkevät liiketoimintapotentiaalia datan keruussa ja analysoinnissa, ja tutkimus- ja kehitystahot (esim. yliopistot ja tutkimuslaitokset) pyrkivät tuomaan esiin massadatan kriittisiä tulevaisuuden kysymyksiä sekä käytännöllisestä että teoreettisesta näkökulmasta. Eri toimijoiden roolien rajat nähdään siis selkeinä. Yhteistä eri toimijoille on kuitenkin se, että tietovarantoja ei nykyisellään hyödynnetä täysimääräisesti. Tutkimukset ovat nostaneet esiin sekä massadatan liiketoiminta- että tuotekehityspotentiaalin, mutta suurin osa kaupungeista ja yrityksistä ei ole tarttunut potentiaaliin.

Miksi massadataa ei siis hyödynnetä? Suurten datamassojen keruuseen ja hyödyntämiseen liittyy luonnollisesti riskejä. Haastatellut kaupungit ja yritykset tunnistivat riskit, mutta tunnustivat sa-

²³ Kaupungeilla viitataan tässä raportissa pääkaupunkiseutuun, Tampereeseen, Turkuun ja Ouluun, joista aineisto on kerätty.

²⁴ Ks. tämän raportin luku 2.

mallalla riskien olevan potentiaalia vaikeammin ennustettavia. Datan käytön lisäämisestä syntyy riskejä sekä yritystoiminnalle (datan väärinkäyttö, datan ostajan kysynnän muutokset), yksityisille henkilöille (datan omistajuuden hämärtyminen tai jakaantuminen, yksityisyydensuojan puute) että julkiselle toimijalle. Kaupallisesta näkökulmasta datan käytön haasteena on myös se, että kuluttajat eivät ole valmiita maksamaan palveluista.

Datan avoimuus herättää sekä kiinnostusta että huolta. Julkisella sektorilla datamassojen käsittelyn suurimpana riskinä pidetään haastattelujen mukaan henkilötietojen suojan murtumista. Murtumista voidaan kuitenkin estää mm. anonymisoidulla dataa. Kaupunkien massadatan tai datan avaamisen asiantuntijat pitävät massadatan avaamista tärkeänä, sillä suuri osa kaupungeissa tuotetusta massadatasta ei kuitenkaan sisällä henkilötietoja. Datalähteiden avaaminen vaatii myös yhteistyötä ja suurta mittakaavaa, sillä vain murto-osan datalähteistä avaaminen ei vielä synnytä mahdollisuutta kehittää kaupungin palveluita tai houkutella yksityisiä toimijoita luomaan uutta liiketoimintaa. Kaupungeissa pidetään myös tärkeänä, että datan avaamisen periaatteet ja tarpeet ovat koko kaupunkiorganisaation tiedossa. Haastatteluissa ja työpajoissa nähtiin ylipäätään olennaisena kiinnittää huomiota datan omistajuus- ja pääsykysymyksiin. Huolta herätti mm. tiedon väärinkäyttö, katoaminen ja tiedon antajien oikeudet. Erityisesti julkisen sektorin toimijoiden tulisi haastattelujen mukaan pyrkiä ”leikkaamaan huhuilta siivet” eli tuoda julkisuuteen tiedon avoimuuteen liittyviä kysymyksiä ja huolenaiheita.

Kaupungit ja kunnat ovat uusien haasteiden edessä. Massadatan kysyntä mm. julkiselta sektorilta tulee kasvamaan tulevaisuudessa. Esimerkiksi sovelluskehittäjät ovat yhä kiinnostuneempia käyttämään kaupungeissa ja kunnissa syntyvää avointa dataa. Haastatellut kaupungit ovat datan avoimuuden, määrän ja laadun suhteen hyvin erilaisia. Yhteistä kaupungeille, ja kuntaorganisaatioille ylipäätään, on datajärjestelmien suhteellinen kankeus ja dataosaamisen puute. Tulevaisuudessa kaupungeilla on oltava yhä enemmän osaamista mm. datan metadatan eli tausta- ja käytettävyystietojen kuvaamisessa. Erityisesti julkisista toiminnoista tai julkisissa tiloissa syntyvän datan on oltava sellaista, että siitä on hyötyä muillekin kuin järjestelmätuottajalle. Kaupunkien tulee huolehtia datan käyttöoikeuksista ja rajapintojen avaamisesta dataa tai datapalveluita ostettaessa.

4.3 Liikenne ja liikkuminen massadatan hyödyntäjänä

Liikenne on yksi massadatan hyödyntämisen painopistealueista²⁵: Älyliikenteen palvelutuotannossa tarvittavaa raakadataa tuottavat kuluttajat, yritykset ja julkinen hallinto. Dataa varastoidaan erilaisiin tietokantoihin. Digitaaliseen tietoon perustuvien liikenteen palvelujen ja sovellusten on arvioitu olevan maailmanlaajuisesti yksi tulevan vuosikymmenen kiinnostavimmista markkinoista. Suomella on mahdollisuus profiloitua liikenteen kokeiluypäristönä. Tätä puoltaa muun muassa Suomen suhteellisen pienimuotoinen liikenne, jolloin pienempiä kokeiluja voidaan hyvin tehdä, usein myös normaalin liikenteen seassa. Parhaimmillaan erilaiset liikenteeseen liittyvät big data -innovaatiot, kuten esimerkiksi liikenneanalyysi ja -ennusteet, voivat olla merkittäviä vientituotteita. Tutkimus- ja innovaatiotukea sekä muuta rahoitusta on liikenteen alalla saatavissa suhteellisen runsaasti. Näin ollen mahdollisuus kansainvälisestäkin merkittävään toimintaan on olemassa.²⁶

Massadataa käytetään liikenteessä sekä ihmisten että koneiden ja laitteiden päätöksenteon tukena. Liikennepuolella dataa voidaan kerätä esimerkiksi liikkumisen dynamiikasta: milloin, miten kauan, minne saakka ja miksi liikutaan. Liikenneverkon kehittämiseen (mm. liikenteen tilannekuva, onnettomuustilanteet ja ruuhkautuminen) voidaan käyttää mm. autojen GPS-tietoa (tulevaisuudessa nelikoptereita), IoT-kulkuneuvoja ja liikennevaloja. Tulevaisuudessa kysymyksiä voi herättää automaattilaitteiden, kuten automaattiautojen tekemät itsenäiset dataratkaisut tai -päätökset.

Liikenteen alalle on luotu viime vuosina paljon erilaisia reaaliaikaiseen ja staattiseen dataan perustuvia palveluita (mm. informaatioon, joukkoliikenteeseen, liikennevaloihin, pysäköinninopastukseen, liikennetiedottamiseen, rekisteritietojen käsittelyyn ja kamerajärjestelmiin). Digitalisoituminen

²⁵ LVM (2014).

²⁶ LVM (2014), s. 39.

tarjoaa mahdollisuuksia kehittää uudenlaisia saavutettavuutta korostavia liikennepalveluja. Siirtämällä painopistettä kohti ekosysteemiajattelua, siilojärjestelmästä yhteistoimintaan, pystytään minimoimaan kehittämisen ja eri käyttäjien kokemat kulkumuodosta toiseen siirtymisen välille jäävät sudenkuopat.

Massadatan hyödyntämisessä olennaisena pidetään datan käytettävyyttä. Esimerkiksi sensorien avulla kerätyn datan tulee olla anonymisoinnin jälkeen käytettävissä mahdollisimman laajasti, mutta ainakin toimijaverkoston kesken. Massadatan käytön ja omistajuuden kysymykset verkottuvat keskenään. Jos esimerkiksi taksi luovuttaa dataa operaattorille tai kaupungille, liiketoimintamallin mukaisesti sen tulee saada jotakin kompensatioksi (esim. rahaa tai käyttöoikeuden dataan).

Liikkuminen palveluna -konseptiin (*Mobility as a Service, MaaS*) liittyvät matkaketjupalvelut tuovat ihmiset liikennejärjestelmän keskipisteeeksi ja niiden rakentaminen tarjoaa yrityksille uusia liiketoimintamahdollisuuksia. MaaS-konseptissa käyttäjille tarjotaan erilaisia liikkumispalveluja samojen käyttäjärajapintojen välityksellä. Tällöin esimerkiksi nykyiset joukkoliikennepalvelut (lähi- ja kaukoliikenne sekä kansainväliset yhteydet), vuokra-autot, kaupunkipyörät, kyytipalvelut, yhteiskäyttöauto-palvelut, taksipalvelut ja useat muut liikennemuodot tulisivat käyttöön yhden käyttöliittymän taakse.

Jotta ihminen tulee aidosti liikkumispalveluiden keskiöön, on noudatettava omadatan periaatteita, joita ovat mm. ihmiskeskeisyys, datan käytettävyys, läpinäkyvä suhteiden hallinta ja avoin liiketoimintaympäristö. Näiden periaatteiden toteutuminen vaatii poikkihallinnollisen toimijan, dataoperaattorin, syntymistä. Käyttöliittymän hallinnoijana tai operoijana voi olla joko yksityinen tai julkinen taho. Massadatan hallinnoinnissa toimivimpana teknologisenä neutraalina alustana pidettiin haastattelujen mukaan julkista tahoja. Toisaalta yksityistä toimijaa pidettiin tehokkaampana ja joustavampana operaattorina ja alustan tarjoajana. Luotetun dataoperaattorin tai -operaattoreiden roolin määrittely on yksi tärkeimpiä omadataekosysteemin haasteita. Myös liikkumisdataa välittävä taho voisi olla joko julkinen tai yksityinen. Esimerkkinä yksityisestä toimijasta on Soneran Reissu-palvelun tammikuussa 2016 alkanut pilotti Hämeenlinnan seudulla, joka tuo kuluttajille sovelluksen juna- ja taksimatkojen tilaamiseen ja maksamiseen.

4.4 Hyvinvointipalvelut massadatan hyödyntäjinä

Hyvinvointipalvelut ovat myös merkittävä massadata-strategian painopiste: Suomesta löytyy maailmanlaajuisesti katsoen poikkeuksellisen laadukkaita ja kattavia terveystietokantoja, muun muassa geenitietoihin liittyen. Tämän vahvuuden vastuullinen hyödyntäminen massadataratkaisuissa voisi tarjota Suomelle merkittävän, globaalin kilpailukykytekijän. Suomessa olisi innovaatiotilausta erityisesti ikääntyvälle väestönosalle suunnatuille ennaltaehkäisevän terveydenhuollon ja itsehoidon palveluille.

Terveydenhuollon kenttä liittyy massadataan lukuisilla eri tavoilla ja Suomessa on alan osaamista esimerkiksi bioinformatiikan ja molekyylieläketieteen alalla. Toinen tärkeä alue liittyy kansalaisten omatoimiseen terveydenhoitoon ja hyvinvointiin liittyvien itsemittausteknologioiden kehitykseen, jossa suomalaisyrityksissä on osaamista kymmenien vuosien ajalta. Terveysdatan laajemman saatavuuden ennustetaan muuttavan ihmisten käsitystä itsestään ja omasta terveydestään. Elämäntavoista ja geenitiedosta kertyvän yhdistetyn massadatan avulla voidaan mahdollisesti tulevaisuudessa ehkäistä tai viivyttää sairastumista ja edistää terveyttä.

Nykyisessä taloudellisessa tilanteessa, varsinkin kuntien osalta, on säästöjen saavuttaminen uuden teknologian avulla mahdollisimman hyvä palvelutaso säilyttäen avainasemassa. Erilaisia dataan perustuvien omahoitomenetelmien käyttöönottoa tulee samalla tukea. On myös tärkeää, että suomalaiset tuntevat kansainvälisen terveysalan massadatakentän ja ovat mukana vaikuttamassa sen kehitykseen. (LVM, 2014, s. 35)

Massadatalta odotetaan terveydenhuollon laadun parantumista ja säästöjä. Data-analyysin odotetaan kehittyvän tavalla, joka auttaa yksittäisen kansalaisen terveysriskien ennustamisessa erilaisten datavirtojen avulla sekä sairauksien ehkäisyssä. Terveystietojen hyötykäyttöön liittyy kuitenkin yksityisyyteen sekä sosiaalisiin ja eettisiin näkökulmiin liittyviä kysymyksiä, jotka on otettava vakavasti. Ilman näiden tekijöiden hyödyntämistä suotuisat kehityskulut tuskin ovat mahdollisia.

Yksi tapa on määritellä vastuullisen aineiston käytön puitteita yhteistyössä eri alojen osaajien kanssa. Suomessa olisikin muun muassa Sote-uudistuksen myötä oiva tilaisuus pohtia tarkemmin, ko-

keilla ja ottaa käyttöön massadataan perustuvia ratkaisuja. Dataan perustuvista hoitomenetelmistä ja -käytännöistä on saatu merkittäviä tuloksia.

Kaupunkien näkökulmasta sosiaali- ja terveystietojen datan avaamiseen ja käyttämiseen liittyy mm. tietojohdantamisen, ristiinkäyttämisen ja anonymisoinnin ongelmia. Tietojohdantamisen haaste konkretisoituu sosiaali- ja palveluiden tuottamisen vanhoissa rakenteissa, vaikka itse palveluiden sisältö, eli esimerkiksi terveyden edistäminen koetaan aktiiviseksi ja ennakoivaksi toiminnaksi.

Henkilötietoja sisältävien rekisteritietojen ristiinkäyttämisen monimutkaisuus vaikeuttaa palveluiden tietojen saatavuutta ja ennakoivuutta. Massadatan anonymisointi mahdollistaisi sen, että palveluiden tuottamisessa siirryttäisiin reagoinnista ennakointiin. Viranomaisen tehtävänä olisi tällöin joko itse louhia ja käyttää tai luovuttaa data luotettavalle taholle louhittavaksi ja myöhemmin ennakoinnissa hyödynnettäväksi.

Käyttäjän näkökulmasta hyvinvointisektorilla nousussa ovat omaan terveyteen liittyvät, monia datalajeja yhdistelevät palvelut. Potentiaalisessa pilotissa voitaisiin kerätä syömis-, elintapa- ja liikku- mistietoa ja havainnoida ja mitata oman terveydentilan muutoksia esimerkiksi älypuhelimilla tai äly- sormuksilla. Tiedonkeruun ja -analysoinnin omakohtaisuutta vahvistaisi vertaistieto ja tiedon käytön sosiaalisuus esimerkiksi keskustelupalstoilla.

4.5 Avoimesta datasta datastrategiaan

Analysoitaessa julkisten palvelutarjoajien massadatan käyttömahdollisuuksia edellä kuvatun strate- giamallin mukaisesti päähavainto on, että erilaisten datojen käsitteleminen omina kokonaisuuksinaan erillisissä strategioissa on väärä toimintatapa. Avoin data, massadata, reaaliaikainen data, organisaation itsensä tuottama data ja kumppaneiden data ovat kaikki osa datakokonaisuutta. Käytännön toiminta vaatii kokonaisuuden jakamista hallittavan kokosiin osahankkeisiin ja -ohjelmiin, mutta niiden suun- nittelun pohjana olisi oltava koko organisaation kattava datastrategia, missä määritellään periaatteet ja yhteiset käytännöt datan käytölle.

Datan hyödyntämiseen pohjautuvissa innovaatio-ohjelmissa on syytä pyrkiä yhteiseen suunnit- teluun ja synergioihin ohjelmien välillä. Avoin data, massadatastrategia, omadatan edistä- mistoimet, kansallisen palveluarkkitehtuurin kehittämisohjelma ja myös datavetoiset toimialakohtaiset kehittämisohjelmat tekevät kaikki työtä saman asian parissa ja tarvitsevat komponentteja ja ratkaisuja toisiltaan. Massadatan käytön neljä hyötyä (kustannussäästöt, prosessien nopeutuminen, uudet tuotteet tai palvelut sekä parempi päätöksenteko) pätevät datan käyttöön yleisemminkin; hyödyt eivät rajoitu vain massadataan.

Suomi on ollut yksi Euroopan pioneereista julkisen tiedon avaamisessa ja pärjännyt kansain- välisessä vertailussa melko hyvin. Helsinki aloitti Helsinki Region Infoshare -palvelun suunnittelun vuonna 2009, yhtenä Euroopan ensimmäisistä kaupunkiseuduista. Toimet ovat laajentuneet valtion- hallintoon ja sektoritutkimuslaitoksiin. Suurimmat suomalaiskaupungit ovat avaamassa dataa varsin tehokkaasti 6Aika Avoin data ja rajapinnat -hankkeen tukemana. Helpoin data avataan ensimmäisenä, joten viime aikoina maamme tahti datan avaajana on hieman hidastunut. Datan avaamiseen velvoitta- vaa lainsäädäntöä ei ole tehty, mikä voi olla riski tulevaisuudessa.

Avoin data -toiminnassa on keskitytty pääasiassa julkiseen dataan, ja sen tuloksia mitattaessa on korostettu datan liiketoiminnallista käyttöä (yrityssovelluksia ja bisnestä). Nyt kun avoimesta datasta on enemmän kokemuksia eri puolilta maailmaa, fokuksa on syytä laajentaa.

Ensinnäkin, avoin data kattaa vain pienen osan datasta, mikä liittyy kaupungin toimintaan. Kau- punki ei ole rakenne, vaan yhteisö, jonka muodostavat kaupunkilaiset, kaupungissa toimivat yritykset ja muut organisaatiot sekä pienimpänä osana kaupungin virkakoneisto. Julkisen sektorin tuottama data, mikä voidaan avata avoimena datana, on arvokasta ja tärkeää, mutta vain pieni osa kaupunkidatan ko- konaisuudesta. Suurin osa kaupungin datasta on kaupunkiyhteisön tuottamaa dataa. Kaupunkidata olisi hyvä saada hyötykäyttöön laajemmin, jotta kokonaishyödyt olisivat suuremmat.

Toiseksi, avoin data on arvokasta, mutta yksityisen sektorin toteuttama datan hyödyntäminen sovelluksissa tai muussa liiketoiminnassa on vain yksi tapa luoda arvoa datasta. Uudet tuotteet ja palvelut ovat vain yksi datavetoisen toiminnan hyöty. Muut kolme arvonaluontikanavaa – kustannus-

säästöt, prosessien nopeutuminen ja parempi päätöksenteko – riippuvat kaupungin ja muiden datan hyödyntäjien omista toimista.

Kokonaisuutena kaupungit hyödyntävät dataa vielä varsin vähän. Vaikka niillä olisi hallussaan tarvittava data, se ei ole aktiivisesti käytössä vaan hajallaan tai vaikeasti löydettävissä. Kokonaisarkkitehtuurin käyttöönoton taso vaihtelee, mutta on useimmiten heikko. Datan hallinta ja säätely laahaa jäljessä. Organisaatioista puuttuu kulttuuri ja tarvittava osaaminen datan hyödyntämiseen.

Datan kunnolliseen hyödyntämiseen tarvitaan avoimen datan lisäksi muun kaupunkiyhteisön dataa. Avoimesta datasta pitää siirtyä kaupunkidataan, missä eri lähteistä tulevaa dataa voidaan yhdistää ja käyttää tehokkaasti kustannussäästöjen, prosessien nopeuttamisen, uusien palvelujen sekä paremman päätöksenteon perustana. Kaupunkiorganisaatio voi johtaa tämän dataekosysteemin rakentamista.

Lontoon esimerkin mukaan tarvittavan datan voi jakaa pääluokkiin esimerkiksi seuraavasti²⁷:

Avoim data: julkisten tai yksityisten tahojen tuottamaa dataa, johon on 1) avattu konekielinen pääsy, 2) jonka lisenssit mahdollistavat datan käytön, uudelleenkäytön ja muokkauksen rajoituksetta ja veloituksetta, ja 3) joka ei sisällä yksityistä tai luottamuksellista tietoa. Avoin data on usein rajattu tarkoittamaan ensisijaisesti julkisten tahojen dataa, mutta datan merkityksen ja käyttämisen kannalta datan tuottajataholla ei ole merkitystä.

Yksityinen data: yksityisten tai julkisten tahojen tuottamaa suljettua tai lisensoitua dataa, jonka käyttöä ja julkaisua rajoittaa yksityisyyden suoja.

Kaupallinen data: yksityisten tai julkisten tahojen tuottamaa dataa, jonka julkaisua ja käyttöä rajaavat maksulliset lisenssit.

Sensoridata: yksityisten tahojen, julkisten tahojen tai kaupunkilaisten tuottamaa avointa tai suljettua dataa, mikä on koottu sensoreiden ja muiden tietoa kokoavien laitteiden avulla (esim. mobiililaitteet).

Joukkoistettu data: kaupunkilaisten digitaalisten laitteiden ja sosiaalisen median avulla tuottamaa dataa.

4.5.1 Ehdotus datan hyödyntämisen malliksi kaupungeissa

Suomessa on hyvää osaamista toimivan kaupunkidatan ekosysteemin toteuttamiseksi. Helsinki Region Infoshareen kehitettyä, kuutoskaupungeissa käytössä olevaa avoimen datan prosessia voidaan käyttää pohjana datastrategian toteuttamiselle käytännössä²⁸. Mallin neljä vaihetta (tiedon tuottaminen, tiedon avaaminen, tiedon jakaminen ja tiedon hyödyntäminen) pätevät kaikkeen kaupunkidataan tuottajasta riippumatta. Laajemmin kaupunkidataan päivitettyä prosessi voisi toimia seuraavasti:

1) **Datan tuottaminen.** Tuotantovaiheessa edistetään olemassa olevan datan löytämistä ja avaamista nykyistä laajempaan käyttöön sekä uuden datan tuottamista ottamalla data huomioon kaikissa infrastruktuuri- ja palvelu-uudistusprosesseissa, helpottamalla yritysten datan tuotantoa kaupungeissa (esimerkiksi sensorien avulla) sekä mahdollistamalla joukkoistettua datan tuotantoa. Data-aineistojen vertailtavuus, löydettävyyden ja käytettävyyden edellyttävät yhteistyötä datan tuottajien välillä. Datan tuotanto tukee digitaalista jakelua ottaen huomioon myös tiedon hyödyntäjät.

Datan harmonisoinnissa, standardoinnissa ja data-aineistojen jatkokehittämisessä tehdään yhteistyötä yli organisaatio- ja toimialarajojen sekä muiden kaupunkien että kansallisten ja kansainvälisten harmonisointi- ja standardointitoimijoiden kanssa.

2) **Datan avaaminen.** Kaupunkidatan clearing house -toiminto avustaa ja koordinoi datan tuottamista ja avaamista. Clearing house on datan tuottajien ja käyttäjien yhteistyöväline. Sen tehtäviin kuuluu esimerkiksi avattavan datan priorisointi, lisensoinnin hallinta, datan tuottajien avustaminen datan avaamisessa, laadunvarmistus sekä käyttäjäpalautteen kokoaminen ja välittäminen datan tuottajille ja datan uudelleen- ja ristiinkäytön toimintamallien kehittäminen.

27 Greater London Authority: London Data Strategy, 03/2016: <http://data.london.gov.uk/dataset/data-for-london-a-city-data-strategy>

28 Helsinki Region Infoshare: <http://www.hri.fi/fi/hri-projekti/hankkeen-osa-alueet/>

- 3) **Datan jakaminen.** Clearing house ylläpitää kaupunkidatan markkinapaikkaa, josta data-aineistot ovat helposti löydettävissä, saatavissa ja hyödynnettävissä. Verkkopalvelua voi ajatella eräänlaisena datan hakukoneena, joka ohjaa datanjanoisen oikean datan luokse. Palvelussa kerätään käyttäjiltä suoraa palautetta aineistoista ja niiden käytettävyydestä.
- 4) **Datan hyödyntäminen.** Kaupunkidata on hyödyllistä koko kaupunkiyhteisölle: hallinnolle, yrityksille, yliopistoille, korkeakouluille, tutkimuslaitoksille ja kansalaisille. Hyödyntämisen vaiheessa edistetään datan käyttöä mm. aktiivisella kehittäjäyhteistyöllä, koulutuksella ja valmennuksella, ja parantamalla organisaatioiden datavalmiuksia esimerkiksi kumppanikoodari-toiminnalla.

Kuvio 13 Helsinki Region Infoshare kaupunkidatan avaamisen eturintamassa



Case: London Data Strategy

Lontoo on ollut yksi avoimen datan johtavista kaupungeista maailmassa. London Data Store^a oli jo lanseeraushetkenä vuonna 2010^b satoja datasettejä. Iso-Britannia kokonaisuutena on johtava avoin dataa julkaiseva ja hyödyntävä maa Euroopassa.

Avoin data on luonut kiistattomia menestystarinoita. Lontoossa aloittanut, Greater London Authorityn liikennedatan varaan palvelunsa kehittänyt CityMapper^c on kasvanut maailman suurimmaksi matkasovellukseksi, mikä toimii jo 31 kaupungissa ympäri maailmaa ja lisää palveluun uusia kaupunkia kiihtyvällä tahdilla.

Kokemustensa pohjalta Lontoo pitää kuitenkin välttämättömänä laajentaa datastrategiansa avoimesta datasta kaupunkidataan. Lontoon uuden (3/2016) datastrategian^d perusteluista:

Mild Improvement by gentle Increment: why Open Data is no longer enough

While this city has a fine history of open data, there is a growing recognition that much more can be done to exploit the value of 'city data' – a new definition we want to firmly establish at the centre of this City Data Strategy. Sharing and using city data more effectively can deliver improved public services, more efficient infrastructure, and wider data-led innovation.

Strategiassa määritellään kaupungin datatoiminnolle neljä päätavoitetta, jotka ovat sovellettavissa kaupunkeihin yleisesti:

Kaupungin datatalous on yksinkertainen ja perustuu kaikkien toimijoiden yhteistyöhön. Datatalouden pitää olla ymmärrettävä ja yhtenäinen datan tuottajille (kaupunki, liikelaitokset, yksityinen sektori, kaupunkilaiset). Tähän päästään toteuttamalla ja edistämällä datan jakamista, harmonisointia, yhteentoimivuutta ja avoimia standardeja sekä tekemällä aktiivisesti työtä yksityisyydensuojan ja tietoturvan varmistamiseksi.

Kaupunki lanseeraa julkisen ja yksityisen sektorin yhteisen ”City Data Marketin”. Tämä avaa datan hallinnan pois silloista sekä mahdollistaa datan rinnakkaiskäytön yli organisaatorajojen.

Datalle rakennetaan koordinoitu ja yhtenäinen infrastruktuuri, mikä mahdollistaa tiedonlouhinnan ja uusien arvokkaiden integroitujen palvelujen toteuttamisen. Avain tähän on harmonisoitu ja yhteentoimiva rikas data yli toimialojen.

Liiketoimintamallit, arvooverkot, palautemekanismit, lisensointimallit, datamarkkinat, työkalut ja datan hallinta ovat yhtä tärkeitä kuin tekniset työkalut ja toimintamallit. Niitä kaikkia tarvitaan toimivaan kaupunkidatan kokonaisuuteen.

Myös strategian kumppaniverkosto on määritelty hyvin ja on sovellettavissa muihin kaupunkeihin.

Datastrategian puutteet. Lontoon datastrategia on monilta osin erinomainen ja sisältää uutta ajattelua. Joidenkin esitettyjen toimenpiteiden osalta tilanne Suomessa on parempi kuin Lontoossa; esimerkiksi lisensointia esitetään Lontoossa laajennettavaksi kaikkeen dataan vasta nyt. Suomalaisessa avoin data -liikkeessä lisensointi on ollut tärkeä osa datan julkaisua alusta alkaen. Täällä tunnistettiin ajoissa, että lisensointi on tapa luoda molempia osapuolia hyödyttävä sopimus datan käytöstä sekä taata datalle jonkinasteinen palvelutaso.

Datastrategiassa on myös joitakin ilmeisiä puutteita. Näistä merkittävimpiä ovat epäkonkreettinen ja hahmoton suhde yksityisyyden suojaan, sekä strategian heikko yhteys kaupunkidataekosysteemin kansalliseen ja kansainväliseen kokonaisuuteen.

Yksityinen tieto mainitaan strategiassa, mutta siinä ei esitetä uusia keinoja kansalaisten asiointitapojen tai tiedonsaantimahdollisuuksien edistämiseksi, eikä myöskään kansalaisten tietoturvan parantamiseksi. Tämä on selvä puute. Kappaleessa 3.4 pohditaan blockchain-tekniologioiden mahdollisuuksia tämän haasteen ratkaisemisessa.

^a London Data Store: <http://data.london.gov.uk/>

^b Charles Arthur, The Guardian, 1/2010: <http://www.theguardian.com/technology/2010/jan/06/london-datastore-launch-johnson-mashups>

^c <https://citymapper.com/cities>

^d Greater London Authority: London Data Strategy, 03/2016: <http://data.london.gov.uk/dataset/data-for-london-a-city-data-strategy>

Kansainvälisillä markkinoillasuurin este uudelle liiketoiminnalle ja parhaiden hyötyjen saamiselle kaupunkidatasta on se, että datakokonaisuus kaupunkien välillä on hajanainen johtuen yhteisten toimintamallien ja standardien puutteesta. Digitaalisten palvelujen markkinat ovat tyypillisesti globaalit. Yksi kaupunki, edes Lontoo, ei ole riittävä markkina-alue menestykselle. **Nykyiset älykaupunkimarkkinat eivät toimi.** Seuraavassa kappaleessa pohditaan keinoja vauhdittaa markkinoiden syntymistä.

4.6 Kansainvälinen yhteistyö

Eurooppalainen kaupunkikehittäjien yhteisö ja Euroopan komissio ovat heräämässä tarpeeseen kehittää yhtenäinen älykkäiden kaupunkien markkinapaikka. Suomella on avoinna mahdollisuuksien ikkuna profiloitua tämän markkinapaikan johtavana kehittäjänä. 6Aika-ohjelma on uraauurtava kaupunkien yhteistyöhanke, joka on herättänyt laajaa kansainvälistä kiinnostusta ja ollut inspiraationa Open & Agile Smart Cities -verkostolle.

Älykkäät kaupungit ovat nopeasti kehittyvä kenttä. Seuraavilla nopeilla toimilla voidaan päästä alkuun: i) 6Aika-ohjelman kansainvälistä yhteistyötä ja näkyvyyttä vahvistetaan, ii) suomalaiskaupungit osallistuvat aktiivisesti Open & Agile Smart Cities -verkoston toimintaan.²⁹

Case: Open & Agile Smart Cities -verkosto

Kansainvälinen Open & Agile Smart Cities -verkosto^a luo perustan datapohjaiselle yhteistyölle, jolla edistetään yhteentoimivia ratkaisuja, yritysten innovaatiotoimintaa ja älykkäiden palvelujen syntymistä. Verkosto käynnistyi kaupunkien aloitteesta vuonna 2015, ja siinä on mukana jo 89 kaupunkia 19 maasta. Suomalaiset kuutoskaupungit ovat verkoston perustajajäseniä. Käytännön työtä toteutetaan 6Aika-strategiaan kuuluvan Avoin data -kärkihankkeen kautta. Verkoston toimistoa Brysselissä johtaa Seppo Haataja.

Open & Agile Smart Cities -verkosto kokoaa eri maiden kaupunkiverkostot yhteen vauhdittamaan yhtenäisten standardien ja periaatteiden käyttöä. Yksittäinen kaupunki ei voi liittyä OASC-verkostoon, vaan mukaan liitytään kansallisena kaupunkien osaverkostona. Tällä mahdollistetaan älykkäiden kaupunkisovellusten ja -ratkaisujen kehittäminen kerralla useampaan kaupunkiin siten, että ne ovat yhteentoimivia sekä kaupunkien välillä että niiden sisällä. Toimintaa ohjaa kaupunkien välinen yhdessä tekeminen ja kokeileminen. Tavoitteena on, että kaupungeissa voidaan rakentaa palveluja, jotka pohjautuvat aitojen käyttäjäkokemusten perusteella valittuihin yhteneviin ohjelmointirajapintoihin ja tietomalleihin.

Käytännössä kaupungit sitoutuvat neljään mekanismiin:

- yhtenäisen rajapinnan käyttöönottoon (ensi vaiheessa FIWARE NGSI^b)
- yhteisten datamallien kehittämiseen (ensi vaiheessa Helsingin koordinoimassa CitySDK-hankkeessa^c kehitetyt rajapinnat)
- yhteensopivien datakatalogien toteuttamiseen (ensi vaiheessa CKAN-alusta^d)
- kokeilukulttuuriin, mikä perustuu ratkaisujen nopeaan testaamiseen käytännössä oikeassa ympäristössä.

6Aika-strategian idea on, että kaupungit sopivat yhteisistä standardeista ja periaatteista, jotta sovelluksia voidaan kerralla kehittää useisiin kaupunkeihin. Samaa mallia levitetään Open & Agile Smart Cities -verkostossa.

^a Open & Agile Smart Cities: <http://oascities.org/>

^b FIWARE NGSI core: <http://docs.ngsi9.apiary.io/#>

^c CitySDK: <http://www.citysdk.eu/>

^d CKAN-alusta: <http://ckan.org/>

²⁹ EU-komissio on käynnistänyt uuden Urban Innovative Actions –ohjelman, jonka ensimmäiset neljä rahoitushakua ovat nyt käynnissä. Urban Innovative Actions: <http://www.uia-initiative.eu/en/12-topics-for-a-better-urban-environment>

5 Innovatiivisia ratkaisuja massadatan hyödyntämiseen

Lainsäädäntö asettaa rajoja henkilötietoja sisältävän massadatan käytölle. Tämän luvun tavoitteena on tuottaa tietoa toimenpiteistä ja ratkaisuista, jotka takaavat riittävän tietosuojan ja tietojenkäsittelyn avoimuuden massadataan perustuvissa palveluissa. Kansainväliset esimerkit osoittavat, että muualla on jo olemassa hyviä esimerkkejä siitä, kuinka massadataa voidaan hyödyntää lainmukaisesti ja kuinka avoimuutta voidaan parantaa.

5.1 Esimerkkien valinta ja lyhyt esittely

Esimerkkien valinnan lähtökohtana on ollut innovatiivisuus massadatan hyödyntämisessä ja toimintatavoissa. Nyt kuvattavat kolme esimerkkiä on valittu laajemmasta listasta³⁰ ja esimerkit ovat tiivistysti seuraavat. Esimerkki 1 käsittelee datan uudelleenkäytön työkalua. Kyseessä on ilmainen hollantilainen online-työkalu, jonka avulla on mahdollista arvioida datan uudelleenkäytön lainmukaisuutta. Esimerkki 2 kuvaa henkilötietoja sisältävän datan jakamista luotetussa ympäristössä. Kyseessä on kaksi esimerkkiä Hollannista ja Belgiasta, missä tehdään kokeiluja turvallisten ympäristöjen rakentamiseksi henkilötietoja sisältävän datan käsittelyä varten. Esimerkki 3 tarjoaa innovatiivisen mallin koskien julkista varainkäyttöä ja tietosuojaa. Kyseessä on online-työkalu Sloveniasta, jonka avulla on mahdollista saada tietoa huomattavasta osasta julkista varainkäyttöä.

5.2 Henkilötietoja sisältävän datan uudelleenkäytön työkalu

Henkilötietoja sisältävän datan uudelleenkäytön työkalu on käytännönläheinen online-työkalu. Työkalun avulla arvioidaan sitä, onko tarkoitukseen A kerättyjä henkilötietoja mahdollista käyttää tarkoituksessa B. Työkalussa on kolmen eri kategorian kysymyksiä, joihin vastaamalla työkalu pisteyttää sen, onko henkilötietoja mahdollista käyttää tarkoitukseen B. Pisteytys perustuu viiteen myöhemmin eriteltävään tekijään.

Työkalun on rakentanut Geonovum. Geonovum on hollantilainen osittain julkinen toimija, joka toimii hollantilaisen infrastruktuurista ja ympäristöstä vastaavan ministeriön ohjauksessa. Työkalu on julkaistu tammikuussa 2016 ja se näyttää olevan Euroopassa ainutlaatuinen. Työkalua tullaan päivittämään ja se on saatavilla osoitteessa <http://privacy.locatielab.nl>.

Monet organisaatiot haluavat yhdistää erilaista hallussaan olevaa dataa muiden toimijoiden datan kanssa. Pelko siitä, ettei tietojen yhdistäminen ole lainmukaista – ja erityisesti se, ettei toiminta ole henkilötietolainsäädännön mukaista – estää organisaatioita saamasta parhaan hyödyn niiden hallussa olevasta datasta. Työkalu ratkaisee tämän ongelman, tai ainakin se toimii alkusysäyksenä keskustelulle organisaation sisällä. Työkalu antaa organisaatiolle ensiarvion arviointiprosessissa ja sen avulla voidaan punnita datan käytön hyödyllisyyttä ja henkilötietojen suojaa.

Uusien teknologioiden tarjoamat mahdollisuudet, kuten paikkatietojen hyödyntäminen toimintojen tehostamiseksi, kiinnostavat sekä julkisia että yksityisiä toimijoita. Koska datan käyttöön liittyy

30 Katso liite 2.

riskitekijöitä, toimijoiden on oltava uusien teknologioiden kokeiluissaankin varovaisia. Paikkatiedot halutaan käyttää hyödyksi suuressa mittakaavassa, koska tämä parantaa paikkatietojen arvoa. Toisaalta henkilötietolainsäädäntö saattaa asettaa rajoituksia paikkatietojen käytölle, jolloin syntyy jännite tavoitteen ja lainmukaisuuden välille. Hoitaessaan julkista tehtävää organisaatiot törmäävät yhä useammin tähän jännitteeseen harkitessaan uusien teknisten innovaatioiden mahdollistamia ratkaisuja, joissa liikutaan laillisesti sallitun rajalla tai lain väärällä puolella.

Henkilötietodirektiivin (46/1995/EY) 6 artiklan mukaan, joka on implementoitu Suomeen henkilötietolain (523/1999), henkilötietojen käsittely, joka ei ole yhteensopiva henkilötietojen käsittelyn alkuperäisen tarkoituksen kanssa on lähtökohtaisesti direktiivin vastaista ja kiellettyä. Käyttötarkoitussidonnaisuutta voidaan rajoittaa vain 13 artiklassa määritellyillä perusteilla. Jotkut jäsenmaat, kuten Hollanti, ovat direktiiviä implementoidessaan hyväksyneet poikkeuksen, joka sallii henkilötietoja sisältävän datan uudelleen käsittelyn tietyissä tapauksissa. Lisäksi datan taloudellisesta arvosta ja datan yhdistelyn määrän kasvusta johtuen Tietosuojatyöryhmä on julkaissut lausunnon asiaan liittyen. (Data Protection Working Party, 2013)

5.2.1 Esimerkkien sisältö

Pohjimmiltaan työkalu tukee alustavaa arviota siitä, onko henkilötietoja sisältävän datan uudelleen käyttö toiseen tarkoitukseen sallittua. Lähtökohtana on, että henkilötietoja sisältävän datan alkuperäinen kerääminen on tapahtunut lainmukaisesti. Näin ollen työkalu on hyödyllinen seuraavien ehtojen täytyessä:

- data sisältää henkilötietoja
- data on jo kerätty käyttötarkoituksessa A
- datan on kerännyt julkinen toimija tai yksityinen toimija suorittaessaan julkista tehtävää
- datan alkuperäinen kerääminen on ollut lainmukaista
- datan kerännyt toimija haluaa käyttää dataa lainmukaisesti toiseen käyttötarkoitukseen B.

Työkalua ei voi käyttää sen arvioimiseen, onko datan uusi kerääminen lainmukaista. Työkalulla arvioidaan jo kerätyn datan uudelleen käytön lainmukaisuutta. Hollannin lainsäädännössä datan uudelleen käyttöä arvioidaan henkilötietolain (Wet bescherming persoonsgegevens (WBP)) artiklan 9 pohjalta.

Työkalun kohderyhmä on julkinen henkilörekisterin rekisterinpitäjä, esimerkiksi ministeriö tai kunta. Rekisterinpitäjän organisaatiossa työkalu on suunnattu henkilöille, jotka vastaavat siitä, voiko henkilötietoja sisältävää dataa käyttää uudelleen tarkoitukseen B. Tausta-ajatuksena on, että ohjelmaa käyttävillä henkilöllä on perustiedot henkilötietolainsäädännöstä, mutta heidän ei tarvitse olla juristeja.

Työkalun sisältämiin kysymyksiin vastaaminen voi viedä muutaman tunnin. Tarkoituksena on, että jo työkalun käyttäminen lisää ymmärrystä asiasta ja siihen liittyvistä kysymyksistä. Työkalu antaa käyttäjälleen ensimmäisen näkemyksen datan uudelleen käytön lainmukaisuudesta. Lisäksi työkalun avulla on helpompi käydä organisaatiossa läpi niitä kysymyksiä, jotka ovat olennaisimpia datan uudelleen käytön lainmukaisuutta arvioitaessa. Työkalulla on ymmärrettävästi myös rajoitteita. Koska henkilötietolainsäädännössä on kyse abstraktista asiasta, niin tarkoituksena ei ole, että datan uudelleen käytön arvioinnissa nojattaisiin vain työkalun antamaan pisteytykseen. Ajatuksena on, että prosessin myöhemmissä vaiheissa käytetään apuna asiantuntijaa.

Työkalu koostuu kolmen eri kategorian kysymyksistä, joihin tulee vastata ennen pisteytyksen saamista. Nämä kolme kategoriaa ovat: i) Kahdeksan pudotusvalikkokysymystä, jotka varmistavat sen, että työkalua käytetään oikealla tavalla, ii) Viisi avointa kysymystä, jotka varmistavat sen, että faktat koskien dataa on kirjattu oikein, ja iii) 35 kysymystä uudelleen käytettävyydestä, jotka toimivat annetun pisteytyksen perusteina siitä, onko datan uusi käyttötarkoitus lainmukainen.

Työkalu linkittyy Hollannin henkilötietolain artiklaan 9, jonka mukaan henkilötietojen uusi käyttötarkoitus ei saa olla yhteensopimaton datan alkuperäisen käyttötarkoituksen kanssa. Artiklan 9 toinen kohta sisältää viisi kriteeriä, joilla tätä yhteensopimattomuutta voidaan arvioida. Nämä kriteerit ovat:

- datan uuden käyttötarkoituksen ja datan alkuperäisen käyttötarkoituksen yhtenevyys
- minkälaisesta henkilötietoja sisältävästä datasta on kysymys
- henkilötietojen uuden käsittelyn vaikutus rekisteröityyn
- millä tavalla henkilötietoja sisältävä data on hankittu
- rekisterinpitäjän käyttämät suojausmenetelmät, jotka varmistavat henkilötietojen oikeudenmukaisen käsittelyn ja estävät väärinkäytön.

Työkalu on näiden kriteerien mukainen ja se muuttaa nämä jokseenkin abstraktit kriteerit konkreettisiksi kysymyksiksi. Kun kaikkiin kysymyksiin on vastattu, työkalu pisteyttää tuloksen. Työkalu painottaa määrättyjä kysymyksiä pisteytystä laskettaessa. Jos työkalu antaa tulokseksi alle 50 prosenttia 100 prosentin maksimituloksesta, antaa ohjelma yleisiä ohjeita siitä, kuinka pisteytystä on mahdollista parantaa. Ohjelma voi esimerkiksi ehdottaa suojausmenetelmien parantamista.

5.2.2 Relevanssi Suomessa

Työkalu on käyttökelpoinen Suomessa. Nykyisen henkilötietolain 6 §:n mukaan henkilötietojen käsittelyn tarkoitukset on määriteltävä ennen henkilötietojen keräämistä tai muodostamista henkilörekisteriksi. Henkilötietolain 7 §:ssä todetaan, että henkilötietoja saa käyttää tai muutoin käsitellä vain tavalla, joka ei ole yhteensopimaton 6 §:ssä tarkoitettujen käsittelyn tarkoitusten kanssa. Henkilötietolakia tulkittaessa tulee huomioida se, että laki käyttää ilmaisua yhteensopimaton. Näin ollen rekisterinpitäjällä on jonkin verran vapautta siinä, miten se tulkitsee lain määritelmää. Myöskään henkilötietodirektiivi ei määrittele tiukasti tätä käyttötarkoitussidonnaisuusperiaatetta. Lisäksi henkilötietolaki määrittelee 8 §:ssä henkilötietojen käsittelyn yleiset edellytykset. Henkilötietoja saa käsitellä ainoastaan näiden edellytysten täyttyessä.³¹

Tavallisen henkilön tai yrityksen näkökulmasta tämä voi kuulostaa hyvin monimutkaiselta. Voidaan väittää, että tietotaso henkilötietolain sisällöstä on Suomessa matala. Tämä pahentaa edellä mainittua ongelmaa. Tietosuojatyöryhmä onkin osuvasti todennut, että johdonmukaisen tulkintatavan puuttuminen voi heikentää rekisteröityjen suojaa ja se voi luoda ylimääräisiä sääntelystä johtuvia kustannuksia liike-elämälle ja muille toimijoille. Tästä on vähitellen tullut yhä suurempi ongelma, kun datan volyymi ja saatavuus ovat kasvaneet eksponentiaalisesti. Lisäksi henkilötietojen käsittelystä on tullut yhä olennaisempi osa modernia yhteiskuntaa.³² Yhä kasvavissa määrin massadatan arvo on datan niissä käyttötarkoituksissa, joita rekisterinpitäjä ei tiedä dataa kerätessään.³³ Tämän takia massadatan tehokas käyttö voi olla ristiriidassa käyttötarkoitussidonnaisuuden periaatteen kanssa kuten myös henkilötietolain suhteellisuus- ja tarpeellisuusvaatimuksien kanssa.³⁴ Tulee huomioida, että uusi tietosuojasetus³⁵ tuo raskaampia sanktioita toimijoille, jotka eivät noudata asetusta.

Tämän takia toimijoiden on yhä tärkeämpää tietää se, mitä datalla voi tehdä nykyisessä oikeudellisessa ympäristössä. Henkilötietoja sisältävän datan uudelleenkäytön työkalu mahdollistaa helpon tavan analysoida sitä, voidaanko käyttötarkoitukseen A kerättyä informaatiota käyttää toiseen tarkoitukseen. Tämä on olennainen tieto erityisesti massadatan käytössä. Analysointityökalu olisi arvokas sekä julkiselle että yksityiselle sektorille, minkä takia työkalun mahdollinen käyttö Suomessa ei tulisi rajoittua ainoastaan julkiselle sektorille. Katsottaessa lähitulevaisuuteen uusi tietosuojasetus korostaa työkalun merkitystä entisestään. Uuden tietosuojasetuksen 6 artiklan kohta 3 pitää sisällään lähes

31 Massadatan käyttöön liittyvää sääntelyä käsitellään laajemmin LVM:n (2016) raportissa.

32 Data Protection Working Party (2013).

33 Mayer-Schönberger ja Cukier (2013), s. 12–14.

34 LVM (2016), s. 8.

35 Tätä kirjoittaessa virallista tekstiä ei ole vielä saatavilla. Viittauksemme tietosuojasetukseen koskevat seuraavaa versiota: Brussels, 15 December 2015 (OR. en) 15039/15 LIMITE Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) [first reading] – Analysis of the final compromise text with a view to agreement Interinstitutional File: 2012/0011 (COD) DATAPROTECT 229 JAI 976 MI 786 DIGIT 108 DAPIX 235 FREMP 295 COMIX 663 CODEC 1676

kaikki työkalun elementit ja on pitkälti yhtenevä Hollannin henkilötietolain artiklan 9 kanssa. Tietosuoja-asetuksen myötä yhteensopivuudesta on sovellettavissa myös suomalaisiin rekisterinpitäjiin pitkälti samankaltaisena kuin nyt Hollannissa.

Tämän lisäksi tietosuoja-asetuksen 20 artikla tekee tärkeän uudistuksen koskien profilointia. Tietosuoja-asetuksen 20 artikla on täysin uusi kohta verrattuna tietosuojadirektiiviin ja tämän uuden kohdan tuominen tietosuoja-asetukseen alleviivaa asian merkittävyyttä. Artiklan 20 mukaan luonnollisella henkilöllä on pääsääntöisesti oikeus olla joutumatta sellaisen toimenpiteen kohteeksi, joka aiheuttaa hänelle oikeusvaikutuksia tai vaikuttaa häneen merkittävällä tavalla ja joka on tehty ainoastaan automaattisen tietojenkäsittelyn perusteella. Koska profilointi usein merkitsee aiemmin kerätyn henkilötietoja sisältävän datan uudelleenkäyttöä, niin tietosuoja-asetuksen 20 artiklan tuoman muutoksen myötä työkalun merkittävyys kasvaa tulevaisuudessa.

5.3 Henkilötietoja sisältävän datan jakaminen luotetussa ympäristössä

Julkinen sektori on yhä kasvavassa määrin alkanut ymmärtää hallussaan olevan datan ja erityisesti koottujen tietojen yhdistelemisen arvon. Julkinen sektori voi tehostaa toimintaansa, koska dataa hyödyntämällä voidaan löytää ja ennustaa asioita, joita ei ole aiemmin havaittu. Tällöin näihin asioihin on mahdollista myös reagoida etukäteen. Tämä pätee erityisesti kuntiin, jotka toimivat lähellä kansalaista ja jotka tämän takia tuottavat ja hallitsevat suurta määrää henkilötietoja sisältävää dataa. Kunnat ovat ristiriitaisen paineen alla. Kunnissa ollaan tietoisia henkilötietolainsäädännön asettamista velvollisuuksista, mutta toisaalta mahdollisuus toimia tehokkaammin ja vähemmin kustannuksin on houkutteleva vaihtoehto, vaikka tämä merkitsisi lainsäädännön ja erityisesti henkilötietolainsäädännön asettamien rajojen rikkomista.

Tämän ristiriidan takia joissain Euroopan unionin jäsenmaissa on käynnissä tunnustelut mahdollisuudesta käyttää luotettua kolmatta osapuolta henkilötietoja käsitellessä. Kuvailimme kahta esimerkkiä Hollannista ja Belgiasta tähän liittyen.

5.3.1 Esimerkki 2a – Hollannin tilastoviranomainen

Hollannissa tilastoviranomainen (Centraal Bureau voor Statistiek (CBS)) on arvioimassa uudelleen tehtäviään. Tähän liittyen se tutkii uusia tapoja datan hankkimiseen esimerkiksi ostamalla dataa teleoperaattoreilta, sen sijaan, että se tuottaisi tämän datan itse. Lisäksi CBS arvioi sitä, mikä rooli sillä tulisi olla julkisen datan keräämisessä ja analysoimisessa erityisesti kuntien osalta.

Sekä kunnilla että CBS:llä on selvä oma intressi tässä keskustelussa. CBS:n käsitys on, että yritykset ja julkiset toimijat tulevat valtaamaan perinteisen tilastoviranomaisen liiketoimintamallin, jossa dataa kerätään julkiselta sektorilta, kuten kunnilta, julkaisua varten. Kunnat, jotka ovat velvollisia luovuttamaan CBS:lle laajasti erilaista dataa, haluaisivat saada lisähyötyä tästä luovutetusta datasta. Kunnat haluaisivat käsitellä henkilötietoja sisältävää massadataa tavalla, joka tehostaa niiden toimintaa. Näiden intressien yhteensovittamiseksi CBS on käynnistänyt aloitteen yhteistyössä kuntien, kansallisen tutkimushallituksen (TNO)³⁶ ja Geonovumin kanssa. Tarkoituksena on tunnustella mahdollisuutta CBS:lle uuteen rooliin, jossa se muiden julkisten toimijoiden pyynnöstä yhdistää ja analysoi myös henkilötietoja sisältävää massadataa. Tätä tehdessään CBS:n tulee ymmärrettävästi noudattaa voimassa olevaa henkilötietolainsäädäntöä. Kansallisena tilastoviranomaisena CBS:llä on pitkä kokemus datan käsittelystä tavalla, jolloin tuloksena syntyvistä tilastoista ei voida tunnistaa henkilöitä. Aiotussa viranomaisyhteistyössä CBS toimiikin eräänlaisena luotettuna kumppanina muille julkisen sektorin toimijoille.

Tällä hetkellä aloite on vielä alkutekijöissään, mutta kansallisesti aloite on käynnistänyt vilkkaan keskustelun. Jos aloite viedään eteenpäin, niin olennaista on Hollannin tietosuojaviranomaisen (DPA) suhtautuminen aloitteeseen. Ensinnäkin DPA arvioi tarkasti tällaisia uusia toimintamalleja ja niiden

36 The National Research Board.

lainmukaisuutta. Toisaalta DPA voi nähdä tämän mahdollisuutena ohjata kunnallisia 'villin lännen suunnitelmia', koska monet kunnat tekevät kokeita henkilötietoja sisältävällä datallaan suuressakin mittakaavassa ja nämä kokeilut rikkovat usein henkilötietolainsäädäntöä. Jos CBS ottaisi kuntien roolin henkilötietoja sisältävän datan uudelleenkäsitteilyjänä, niin tämän toiminnan keskittäminen luotettulle viranomaiselle CBS:lle helpottaisi valvontaa.

Esimerkin faktat ovat seuraavat:

- Toteuttaessaan julkista tehtäväänsä kunnat keräävät ja hallitsevat suurta määrää eräänlaista mikrodataa koskien esimerkiksi asumistukia, kunnallisia kuljetuspalveluita vanhoille ja vammaisille ihmisille, sosiaaliturvatukia, jätehuoltoa, eräitä rikostilastoja jne.
- Hollannin lainsäädännön mukaan kunnat ovat velvollisia jakamaan suurimman osan tästä datasta CBS:n kanssa.
- Hollannin lainsäädäntö velvoittaa CBS:n prosessoimaan tämän datan ja tekemään niistä tilastoja, jotka palvelevat yleistä tiedontarvetta.
- Tämä data sisältää suurimmaksi osaksi tietoja, jotka ovat henkilötietoja sekä eurooppaoikeuden että Hollannin lainsäädännön perusteella. Kuntien hallussa olevan datan toimittaminen CBS:lle ja CBS:n tämän datan analysointi katsotaan datan käsittelyksi.

Datan yhdisteleminen ja analysoiminen uusiin käyttötarkoituksiin voi olla laitonta Hollannin henkilötietolainsäädännön näkökulmasta. Kuitenkin kun analysoiminen ja yhdisteleminen tehdään tilastollisiin tarkoituksiin CBS:n toimesta, niin tällöin syntyy eräänlainen suojasatama, jonka puitteissa voidaan toimia laajemmin kuin normaalisti. Lisäksi CBS:n mukanaolo luo toiminnalle lisäuskuuttavuutta. Tästä johtuen vaihtoehto on kiinnostava monille kunnille, jotka haluavat datojen analysointia ja yhdistämistä muihin datoihin.

5.3.2 Esimerkki 2b – belgialaiset henkilötietoja sisältävät datavarastot

Belgiassa Flanderissa on aloitettu erillisten datavarastojen luominen julkisella sektorilla syntyneitä henkilötietoja sisältävää dataa varten. Tarkoituksena on luoda yhtenäiset ohjeet ja käytänteet näille datavarastoille. Tausta tälle datavarastojen luomiselle on se, että Flanderi on viime aikoina pilkkonut isoja julkisen sektorin toimijoita. Tämän seurauksena useat uudet toimijat hallitsevat ja käyttävät henkilötietoja sisältävää dataa omissa datavarastoissaan. Koska uusilta toimijoilta puuttuu usein syvempää tietämystä henkilötietolainsäädännöstä, niin marraskuussa 2015 Flanderin julkisen hallinnon tietoliikenteestä vastaava viranomainen (Vlaamse Toezichtcommissie voor het elektronische bestuursverkeer) julkaisi ohjeet, jotka säätelevät uusien datavarastojen perustamista.

Esimerkin datavarastoa ei voi määritellä osaksi viranomaista tai erilliseksi toimijaksi. Se on enemmänkin erillinen virtuaalinen varasto, johon pääsyä ja käyttötarkoituksia on rajoitettu. Datavarasto on tietovaranto, jonka sisältöä voidaan vain lukea. Viranomaisen yleistä toimintaa harjoittavat tahot eivät voi tuoda tai muokata datavarastoon tallennettua dataa. Toisin sanoen datavarastojen hallinta ja käyttö on erotettu hallinnon normaaleista operatiivisista toiminnoista. Tyypillisesti datavarastoja käytetään hyväksi laajemmassa poliittisessa päätöksenteossa ja arvioinneissa sen sijaan, että datavarastoja käytettäisiin konkreettisten yksittäisten päätösten apuna, joilla olisi vaikutuksia yksityisiin tahoihin.

Yksityisyyden suojaamisen ja datan sisältämän muun salassa pidettävän tiedon suojaamisen varmistamiseksi datavaraston sisältämä data salataan kolmannen osapuolen toimesta. Pääsy datavarastoon on suojattu salasanalla ja henkilöiden määrä, joilla on pääsy datavarastoon, on hyvin rajoitettu. Suojaprosessin noudattaminen ja toimiminen auditoidaan vuosittain, ja lisäksi datavarastossa oleva data on mahdollista anonymisoida³⁷. Datavaraston käyttäjillä ei ole pääsyä suojaamattomaan dataan. Tämän lisäksi luotettu kolmas osapuoli valvoo käytön suhteellisuutta ja sitä, mihin haettua dataa käytetään. Datan anonymisoinnin ja henkilötietojen suojan varmistamiseksi datavarastosta ei voi hakea yksittäisiä tietoja, vaan minimissään otoksessa on oltava vähintään viisi tietuetta. Nämä säännöt ja toimintatavat

³⁷ Anonymisointia käsittelee tarkemmin LVM (2016), s. 18–21.

luovat standardin, jonka avulla varmistetaan henkilötietolainsäädännön noudattaminen, yksityisyyden suojan kunnioittaminen ja datavarastojen lainmukainen hallinnointi.

Luotu standardi pätee datavaraston perustamiseen, käyttöön, ylläpitoon ja lopettamiseen. Ajatuksena on, että datavarastojen sisältämää dataa käytetään hallinnon sisällä syntyvien kiireellisten kysymysten ratkaisemisessa. Kun tarvittava informaatio haetaan datavarastoista, eikä datan tuottaneelta viranomaiselta, haku ei kuormita kyseistä viranomaista ja dataa on mahdollista käyttää itsenäisesti. Datavarastojen ohjelmisto mahdollistaa datan tilastollisen analyysin, mukaan lukien data-aineistojen yhdistelemisen. Tarkoitus on, että datan hankkiminen datavarastoista on nopeaa ja kevyttä.

5.3.3 Esimerkkien relevanssi Suomelle

Esimerkit ovat sovellettavissa Suomeen. Suomalaiset kunnat keräävät ja hallitsevat suurta määrää dataa eri sektoreilta. Myös Suomessa tilastolaki (280/2004) edellyttää, että kunnat sekä muut julkiset ja yksityiset toimijat jakavat tilastolaissa erikseen määritellyn tilastotuotantoa varten tarvittavan datan tilastoviranomaisten kanssa, joista päätoimija on Tilastokeskus.

Jos katsomme asiaa sekä eurooppaoikeuden että Suomen kansallisen henkilötietolain näkökulmasta, niin henkilötietoja sisältävän datan käsittely on sallittua tilastointia varten.³⁸ Näin ollen Suomea velvoittava lainsäädäntö on esimerkissä pitkälti yhtenevä Hollannin lainsäädännön kanssa. Kun toiseen tarkoitukseen kerättyä henkilötietoja sisältävää dataa hyödynnetään tilastolliseen tarkoitukseen, niin tämä käyttö ei ole ristiriidassa henkilötietolain käyttötarkoitussidonnaisuuden periaatteen kanssa.

Kunnat, jotka haluavat hyödyntää hallussaan olevaa massadataa voisivat tehdä toimeksiantosopimuksen Tilastokeskuksen kanssa, jonka perusteella Tilastokeskus laatii halutut tilastot. Näitä tilastoja, joista yksityiset henkilöt eivät voi olla tunnistettavissa, voitaisiin sitten käyttää yhteiskunnallisessa päätöksenteossa. Tämän takia on mielenkiintoista ajatella sitä, miten Tilastokeskuksen rooli voisi kehittyä massadatan tarjoajana ja käsittelijänä. On huomioitava, että tilastolain 2 § velvoittaa Tilastokeskusta ja muita tilastoviranomaisia tarjoamaan tilastotietoa, jota yhteiskunnassa tarvitaan ja jonka on tarpeen olla yleisesti saatavilla.³⁹ Tällöin kuitenkin tulee huomioda Tilastokeskuksen käytettävissä olevat resurssit.

Henkilötietoja sisältävän datan kerääminen tilastotarkoituksiin tulee tapahtua henkilötietolain ja tilastolain mukaisesti. Tilastotarkoituksiin kerättävän tiedon luovuttaminen on vapaaehtoista, ellei luovutuksesta ole säädetty tilastolaissa. Esimerkiksi kunnat ovat velvoitettuja antamaan tilastojen laatimista varten välttämättömät tiedot taloudestaan, hyödykkeistään, toiminnan sijainnista, henkilöstöstään ja muista toiminnan edellyttämistä voimavaroista tilastolain 14 § 2 momentin perusteella. Tilastolain 5 §:n mukaan tiedonantajilta, kuten kunnilta, pyydetään vain tilastojen laatimisen kannalta välttämättömät tiedot. Laissa mainittu välttämättömyysedellytys suojaa tiedonantajien oikeuksia.

Välttämättömyysedellytys datalle, jota kerätään tilastointitarkoituksiin tilastolaissa säädetyn tiedonantovelvollisuuden nojalla, voi olla ongelmallinen massadatan tehokkaassa hyödyntämisessä. Ennen tietojen keräämistä voi olla vaikea arvioida kerätyn tiedon kaikkia hyödyntämismahdollisuuksia, jolloin voi olla vaikea määritellä sitä, mikä on välttämättömä. On myös taloudellisesti mahdotonta, että Tilastokeskus keräisi valtavan määrät laillisesti saatavilla olevaa dataa, josta saattaisi olla massadatana hyötyä tulevaisuudessa. Tämän takia kuntien vapaaehtoisesti toimeksiantosopimuksen nojalla mahdollisesti luovuttama aineisto on keskeistä. Tällöin kyse on sellaisen datan analysoinnista, jota kunnat pitävät tärkeänä ja datan luovuttaminen ei perustu tilastolain tiedonantovelvollisuuteen.

Suomen henkilötietolainsäädännön velvoitteet pätevät myös datavarastoissa oleviin tietoihin ja niiden käsittelyyn. Jos suomalainen viranomais haluaisi perustaa belgialaista datavarastoa vastaavan toimijan, niin sen tulisi lähtökohtaisesti valita, vastaako se itse datavarastosta vai onko datavaraston ylläpitäjä erillinen oikeushenkilö. Jos datavaraston ylläpitäjä olisi erillinen oikeushenkilö, niin tällöin datavaraston ylläpidosta tulisi tehdä toimeksiantosopimus, jossa osapuolten vastuut ja velvollisuudet määriteltäisiin.

38 Henkilötietodirektiivin 6 artiklan, uuden tietosuojasetuksen 5 artiklan ja henkilötietolain 7 § mukaan henkilötietojen käsittelyä tilastotarkoitusta varten ei pidetä yhteensopimattomana alkuperäisten käsittelyn tarkoitusten kanssa.

39 Katso HE 157/2003 vp, sivu 19.

Datan anonymisointia ei tehdä kovin laajamittaisesti suomalaisissa kaupungeissa. Osa kaupungeista on anonymisoinut henkilötietoja sisältävää dataa tutkimuskäyttöön itse. Aineistossa on saattanut olla mukana muiden toimijoiden dataa, mutta salausta on tapahtunut kunkin omasta toimesta.

Kaupungit eivät vielä hyödynnä täysimääräisesti anonymisointipalveluja, joita Suomessa tarjoavat sekä Yhteiskuntatieteellinen tietoaarkisto (FSD) että Tilastokeskus. FSD:llä tarkastetaan alkuperäisen tutkijan jo tekemä anonymisointi ja aineistoa anonymisoidaan tarvittaessa lisää. Yhteiskunnallisella tietoaarkistolla ei kuitenkaan ole resursseja eikä osaamista massadatan anonymisointiin tai arkistointiin. Tilastokeskus tarjoaa maksullista tutkijapalvelua tutkimus- ja muiden aineistojen muokkaamiseen. Datat anonymisointia tehdään muiden toimeksiantojen yhteydessä, eikä pelkästään anonymisointiin liittyviä toimeksiantoja ole tehty. Erään suomalaisen kaupungin tietopalvelun mukaan perustapauksessa henkilöä koskevaa tietoa kootaan eri rekistereistä. Aineistoa voidaan hyödyntää etäyhteyden kautta, eli dataa ei saa järjestelmästä ulos, mutta sen sijaan voidaan tehdä erilaisia tilastollisia analyysejä ja nämä analyysien tulokset sitten on mahdollista (erillisen lupakäsittelyn kautta) liittää myös omiin raportteihin ja julkaisuihin.

Tietoaarkistot ovat itsekkin uusien haasteiden edessä massadatan hyödyntämisessä ja osallistuvat nyt aktiivisesti kotimaisiin ja kansainvälisiin kehitysryhmiin. Esimerkiksi CESSDA (Consortium of European Social Science Data Archives) on mukana Big Data Europe -hankkeessa (<http://cessda.net/CESSDA-Services/Projects/Big-Data-Europe>). Suomessa ATT-hanke (<http://avointiede.fi/>) etsii ratkaisuja tiede- ja tutkimusmaailman osalta. Keskustelua kansallisesta ”anonymisointikeskuksesta” on käyty.

5.4 Avoin julkinen varainkäyttö ja tietosuoja (Supervisor)

Kolmas alue, jossa on näkyvissä merkittävä muutos myös henkilötietojen käsittelyssä, on hankintojen avoimuus julkisella sektorilla.⁴⁰ Avoimuuden painoarvo on kasvussa, ja tämä on näkynyt useissa aloitteissa ympäri Eurooppaa.⁴¹ Supervisor on hieno esimerkki Sloveniasta avoimuuden lisäämisestä julkisissa hankinnoissa.

Julkisilla hankinnoilla on valtava merkitys Euroopan taloudelle. Euroopan komission tekemän arvion mukaan julkisten hankintojen arvo oli vuonna 2011 noin 2,4 biljoonaa euroa. Tämä on 19 % Euroopan bruttokansantuotteesta. Suomessa julkiset hankinnat kattavat noin 17 % Suomen bruttokansantuotteesta. Koska julkisissa hankinnoissa kyse on veronmaksajien rahoista, on tarve varmistua siitä, että julkiset varat käytetään tehokkaimmalla mahdollisella tavalla ja hankintoihin osallistuvilla yrityksillä on mahdollisuus kilpailla tasaväkisesti. Hankintoihin, jotka ylittävät hankintadirektiiveissä säädetyt kynnysarvot, sovelletaan eurooppaoikeudellista hankintalainsäädäntöä.

Joutuessaan toimimaan aiempina vuosikymmeninä syntyneiden perinteiden ja toimintatapojen kanssa, Slovenian korruptionvastaiseen toimintaan keskittynyt viranomainen (Slovenian Commission for the Prevention of Corruption CPC) keräsi ja analysoi Slovenian rahan liikkeitä julkisen ja yksityisen sektorin välillä vuosina 2003–2010. Tuloksista kävi ilmi, että vain pienelle yritysjoukolla suunnatut maksut lisääntyivät huomattavasti hallituksen vaihtumisien yhteydessä. Kävi myös ilmi, että kilpailu oli vähäistä ja markkinat olivat monopoliset tietyillä talouden sektoreilla (it-palvelut, lääketeollisuus ja rakennuspalvelut). Eräät hallinnonalat (viestintä ja oikeudelliset palvelut) käyttivät huomattavasti ulkoisia palveluntarjoajia, vaikka kyseisillä hallinnonaloilla olisi tullut olla omat riittävät resurssit tehtäviensä hoitamiseen.

⁴⁰ Keskeisenä periaatteena julkisissa hankinnoissa on avoimuus. Avoimuus on kirjattu koko julkista hankintaa kuvaavaksi tavoite- ja menettelyperiaatteeksi sekä aiemmassa hankintadirektiivissä (2004/18/EY) että uudistetussa hankintadirektiivissä (2014/24/EU).

⁴¹ Supervisorin kanssa samansuuntaisia aloitteita ovat BASE-portaali Portugalissa (www.base.gov.pt), joka on kansallinen internet-portaali, joka tarjoaa keskitetysti tietoa julkisen hallinnon tekemisistä sopimuksista. Toinen samansuuntainen esimerkki on Kroatian julkisten hankintojen tietokanta (<http://integrityobservers.eu>). Tietokanta kokoaa julkisten hankintojen täytäntöönpanoon liittyvää tietoa, mukaan lukien tietoa julkisiin hankintoihin osallistuvista yrityksistä.

5.4.1 Esimerkin sisältö

Tässä kontekstissa CPC yhdessä Slovenian valtiovarainministeriön, julkisia maksuja hoitavan viraston ja rekisteriviranomaisen kanssa käynnisti Supervisor-projektin. Supervisor on online-palvelu, joka kerää ja julkaisee dataa kolmesta lähteestä:

1. Julkisia maksuja hoitava virasto huolehtii suorien ja epäsuorien budjettikäyttäjien maksuliikenteestä, ja tämä data tallennetaan yhteen järjestelmään.
2. Slovenian rekisteriviranomainen tallentaa ja jakaa kaupparekisteritiedot slovenialaisista yrityksistä ja muista yhteisöistä, mukaan lukien slovenialaisten yritysten tilinpäätöstiedot.
3. Virallinen lehti ja Slovenian valtiovarainministeriö antavat tiedon julkisista hankinnoista.

Näistä lähteistä saatuun dataan on mahdollista tutustua ja tehdä hakuja internet-selaimessa toimivassa sovelluksessa. Supervisor tekee avoimeksi sen, kuinka Slovenian julkinen sektori käyttää noin 4,7 miljardia euroa julkisiin hankintoihin vuosittain. Supervisorista on mahdollista saada tietoa sopimusosapuolista, varojen suurimmista saajista, tehtyjen rahasiirtojen päivämääristä ja summista sekä varojen siirtojen perusteista. Nämä tiedot koskevat yli 4 000 euroa ylittäviä varainsiirtoja. Supervisorista on mahdollista saada kuvaajia kyseisistä asioista ja tehdyt haut on mahdollista kohdistaa haluttuun ajankohtaan.

Supervisor oli välitön menestys. Julkaisuhetkellä elokuussa 2011 Supervisor keräsi kuudessa päivässä 2 216 000 vierailua. Määrä on huomattava, koska Sloveniassa on noin kaksi miljoonaa asukasta. Aluksi vierailujen syy oli uteliaisuus. Tällä hetkellä informaatiota käytetään kuitenkin analyttisemmin ja palvelun merkittävimmät käyttäjät ovat tutkivat journalistit ja valtion viranomaiset, jotka tutkivat talousrikoksia, julkista varainkäyttöä ja korruptiota.

Ennen Supervisoria osittaista tietoa varainsiirroista julkisen ja yksityisen sektorin välillä oli jo mahdollista saada perustuen Slovenian julkisuuslakiin. Supervisor kuitenkin teki näistä tiedoista helpommin saatavilla olevia. Jotta Supervisor olisi voimassa olevan slovenialaisen henkilötietolainsäädännön ja julkisuuslainsäädännön kanssa sopusoinnussa, niin CPC loi Supervisoriin tietokonealgoritmin, joka poistaa näytettävistä tiedoista henkilötiedot ja muut lain mukaan salassapidettävät tiedot. Poistettu data sisältää esimerkiksi tietoa yksityisille henkilöille tehdyistä varainsiirroista: palkoista, kulukorvauksista, veronpalautuksista jne.

Henkilöyhtiöitä koskeva data, kuten yhtiön nimi ja osoite, jotka usein ovat samat kuin toimintamiyrittäjän omat tiedot, julkaistaan kokonaisuudessaan Supervisorissa. Näistä tiedoista on mahdollista tehdä hakuja palvelussa. Toisin sanoen avoimuus on olennaisempi arvo kuin näiden henkilöiden henkilötietojen suoja. Julkaistessaan näitä tietoja CPC pystyi tukeutumaan voimassa olevaan lainsäädäntöön. Slovenian korruption estämiseen pyrkivä laki⁴² antoi CPC:lle oikeuden julkaista näitä tietoja suoraan ilman asianomaisen suostumusta.

Mielenkiintoista on, että tietyt huomattavia varoja vastaanottaneet yksityishenkilöt pysyivät Slovenian tietosuojavaltuutetulta kannanottoa kyseisen toiminnan lainmukaisuudesta. Tietosuojavaltuutettu totesi, että suurimmassa osassa tietojen julkaiseminen oli lainmukaista. Tietosuojavaltuutettu kiinnitti huomiota siihen, että CPC tavoitteli toiminnallaan avoimuuden lisäämistä ja se julkaisi henkilötietoja vain tilanteessa, jossa yksityinen henkilö oli vastaanottanut enemmän kuin 150 000 euroa vuosikymmenessä. Slovenian lainsäädäntö salli CPC:n kerätä dataa, mukaan lukien henkilötietoja sisältävää dataa, jo olemassa olevista viranomaisten tietolähteistä. CPC:n ei tarvinnut kerätä uutta dataa tai tehdä raskaita uusia tietoteknisiä sovelluksia Supervisoria varten.

5.4.2 Relevanssi Suomelle

Ensi katsomalla Supervisor ei vaikuta relevantilta Suomelle. Viimeisimmässä Transparency Internationalin julkaisemassa korruptioindeksissä Suomi oli maailman toiseksi vähiten korruptoitunut maa.⁴³ Kuitenkin tuoreessa poliisiammattikorkeakoulun tekemässä korruptiotutkimuksessa julkiset hankin-

⁴² Integrity and Corruption Prevention Act.

⁴³ Katso <https://www.transparency.org/country/#FIN>

nat olivat yhtenä viidestä riskikohteesta korruptiolle Suomessa.⁴⁴ Lisäksi Euroopan komission vuonna 2013 tekemässä tutkimuksessa 25 prosenttia kolmen edellisen vuoden aikana julkisiin hankintoihin osallistuneista suomalaisista yrityksistä katsoi, että korruptio on estänyt heitä voittamasta julkista hankintaa.⁴⁵

Supervizorin kaltainen online-palvelu lisäisi julkisten hankintojen avoimuutta Suomessa. Julkisten hankintojen avoimuuden lisäämisellä on mahdollista tehostaa julkisten hankintojen markkinoita. Esimerkiksi avoimuutta lisäävät toimet toisessa jäsenvaltiossa Slovakiassa aiheuttivat sen, että keskimääräiset tarjoukset per hankinta kasvoivat viidessä vuodessa 2,3 tarjouksesta per tarjouspyyntö 3,6 tarjoukseen per tarjouspyyntö.⁴⁶ Lisäksi tulee huomioida, että Suomi saa verrattain huonoja pisteitä sen julkisen datan avoimuudesta, jota juuri Supervizor tekee julkiseksi. Open Data Indexin mukaan Suomi saa 45 % mahdollisesta 100 %, kun kyse on julkisten hankintojen avoimuudesta. Tämän lisäksi kaupparekisteritietojen ja julkisen varainkäytön julkisuudesta Suomi saa vain 5 %.⁴⁷

Nykyisen lainsäädännön ja oikeuskäytännön valossa hankintasopimusten hintatietojen yleisöjulkisuus voi olla ongelmallista. Julkisuuslaki (1060/2000) toteaa, että salassapidettävää ovat yksityiset liike- ja ammattisalaisuudet, jollei erikseen toisin säädetä (JulkL 20 § 1 momentti 20 kohta). Tämän lisäksi julkisuuslakia sovelletaan lain 4 §:ssä määriteltyihin viranomaisiin⁴⁸. Korkein hallinto-oikeus on todennut ratkaisussaan (2007:83), että julkisten hankintojen tarjousasiakirjojen sisältämät yksikköhinnat ovat tarjouksen tehneen elinkeinonharjoittajan liikesalaisuuksia. Näin ollen nämä tiedot eivät ole yleisöjulkisia. Tulee kuitenkin huomioida, että asianosaisella, kuten toisella tarjoajalla, on oikeus saada tieto tarjousten vertailussa käytetystä hinnasta tai muusta tekijästä (JulkL 11 § 1 momentti 6 kohta).

Jos avoimuuden lisäämistä Supervizorin kaltaisen palvelun avulla halutaan edistää, tälle voi olla mahdollisuus huolimatta edellä mainituista julkisuuslain tulkinnoista. Tyypillisesti julkisten hankintojen tarjouspyynnöissä tarjoajia pyydetään merkitsemään tai nimeämään ne tarjouksen osat, jotka sisältävät liike- ja ammattisalaisuuksia. Jos tarjouksen hintaa ei ole merkitty liike- ja ammattisalaisuudeksi hankintayksikön selkeästä ja yksiselitteisestä kehotuksesta huolimatta, kyse ei ole tarjoajan tulkinnan mukaan liike- ja ammattisalaisuudesta ja tieto voidaan saattaa yleisöjulkiseksi. Tämän lisäksi hankintayksikkö voi lisätä hankintasopimukseen sovellettaviin vakioehtoihin lauseen, jonka mukaan tarjouksen tehnyt toimija sitoutuu siihen, että tarjouksen hinnan saa julkaista yleisölle. Tällainen lisäys on mahdollista saada laajasti käytetyksi lisäämällä se julkisissa hankinnoissa käytettyihin vakioehtoihin tai edellyttämällä eri hankintaohjesäännöissä tällaisen vakioehdon käyttöä. Vaikka Suomessa ei ole olemassa Slovenian tapaista keskitettyä julkisten maksatusten järjestelmää, niin on mahdollista luoda ohjelma, mihin voittaneiden tarjoajien tarjoushinnat on mahdollista syöttää. Tämä informaatio voidaan yhdistää Hilmassa jo valmiina olevaan muuhun dataan, joka koskee kyseistä tarjousta.

5.5 Päätelemät laajemmassa kontekstissa

Teknologian kehitys merkitsee aiemmin tuntemattomia mahdollisuuksia luoda, varastoida, analysoida ja jakaa valtavat määrät dataa silmärepyksessä ja minimaalisin kustannuksin. Tämä muutos tulee vaikuttamaan valtion, kansalaisten ja yritysten rooleihin. Syvimmin tämä muutos vaikuttaa tietosuojaan. Läpikäydyt esimerkit osoittavat seuraavaa:

1. Pyrittäessä sekä hyödyntämään massadatan koko potentiaalia että noudattamaan tietosuojalainsäädäntöä, innovatiivisia yhteisiä ratkaisuja voidaan löytää. Tästä ovat esimerkkinä tapaukset hollantilaisesta tilastoviranomaisesta ja belgialaisista datavarastoista.
2. Toinen trendi on sellaisten työkalujen ja standardien kehittäminen, jotka auttavat ymmärtämään hyvinkin abstraktia ja monimutkaista tietosuojalainsäädäntöä, kuten hollantilainen

44 Peuralainen, ja Mutttilainen (2015).

45 TNS Political & Social (2014), s. 54.

46 Furnas (2013).

47 Katso <http://index.okfn.org/place/finland/>

48 Hankintalain mukaisesti hankintayksiköihin ei sovelleta julkisuuslakia, mikäli nämä eivät ole julkisuuslaissa määriteltyjä viranomaisia.

henkilötietoja sisältävän datan uudelleenkäytön työkalu osoittaa. Tällaisissa työkaluissa on huomattava potentiaali, koska ne yksinkertaistavat abstraktia lainsäädäntöä ja käsitteitä helpommin ymmärrettävään muotoon. Hyöty kasvaa, mikäli tietosuojaviranomainen valvoo työkaluja ja niiden laatua. Esimerkkityökalu voi lisäksi helpottaa ja automatisoida tietosuojaviranomaisten omaa työtä.

3. Kolmanneksi tietosuoja ei ole absoluuttinen oikeus, kuten esimerkki Supervisor osoittaa. Josain tilanteissa periaatepunninta tietosuojan ja muiden julkisten intressien välillä on tarpeen, kun otetaan huomioon yhteiskunnalle syntyvät hyödyt.

Lopuksi on tärkeää ymmärtää, että massadatan ja avoimen datan aikaansaama keskustelu tietosuojasta, on vain yksi oikeudellinen sektori, jota oikeudellinen muutos koskettaa. Massadata tulee herättämään kysymyksiä myös seuraavilla oikeuden aloilla:

Kilpailulainsäädäntö. Suuret datamäärät aiheuttavat sen, että kilpailulainsäädännössä tulee ajatella perinteisten alueeseen ja tuotekategorioihin perustuvien markkinoiden määritelmistä poikkeavilla tavoilla.

Tekijänoikeudet. Jossain maissa on jo esitetty vaatimuksia tekijänoikeuksien vahvistamiselle, mikä toimisi kannustimena laajoja datamassoja hallitseville toimijoille kaupallistaa massadataa, jolloin yhteiskunta voisi hyötyä laajemmin datanvarannoista. Toisaalta mahdollisuus luoda massadataa suurissa määrin ja nopeasti aiheuttaa sen, että esimerkiksi aiemmin työläitä tietokantoja voidaan luoda ilman huomattavaa taloudellista tai muuta panostusta.

Yksityisoikeus. Kun eri lähteistä olevaa dataa yhdistellään silmänräpäyksessä, niin kysymykset datan hallinnasta, vastuista ja julkisesta vastuusta sekoittuvat.

Julkisoikeus. Mistä julkisen vallan vastuut ja velvoitteet alkavat ja loppuvat? Mitä tulee jättää markkinoiden hoidettavaksi ja mitä ei? Toisin sanoen massadata pakottaa julkisen vallan uudelleen arvioimaan tehtäviensä. Esimerkiksi miltä osin massadata tulee nähdä julkisena hyödykkeenä?

5.6 Suositukset: esimerkkityökalujen sovellettavuus Suomeen

Ensimmäinen esimerkki, henkilötietoja sisältävän datan uudelleenkäytön työkalu, on sovellettavissa Suomessa. Työkalu konkretisoi abstraktia henkilötietolainsäädäntöä ja automatisoi henkilötietoja sisältävän datan uudelleenkäytön lainmukaisuuden arviointia. Massadatan hyödyntämisessä jo kerätyn datan uudelleenkäyttö on olennaista, mikä korostaa työkalun merkitystä. Mikäli työkalua tarjotaan sekä yksityisille että julkisille toimijoille, työkalun käyttäjäkunta laajenee alkuperäisestä esimerkistä. Parhaimmillaan työkalu voi pienentää tietosuojavaltuutetun työtaakkaa ja työkalu on jo itsessään tapa informoida henkilötietolainsäädännöstä ja sen velvoitteista.

Toinen esimerkki, Hollannin tilastoviranomainen datan uudelleenkäyttäjänä, on mielenkiintoinen, koska henkilötietolaissa edellytettyä käyttötarkoitussidonnaisuutta tulkitaan lainmukaisella tavalla. Tämä laajentaa julkisien toimijoiden mahdollisuuksia hyödyntää heidän hallussaan olevaa henkilötietoja sisältävää massadataa. Mikäli massadatan merkitys kasvaa ennusteiden mukaisesti tulevaisuudessa, myös Suomessa voi olla tarpeellista käydä sama keskustelu tilastoviranomaisen roolista massadatan uudelleenkäyttäjänä kuin mitä Hollannissa jo käydään.

Supervisor on hieno esimerkki siitä, kuinka julkisen vallan koko varainkäyttö on avoimena ja helposti saatavilla. Kun otetaan huomioon julkisissa hankinnoissa liikkuvat valtavat rahavirrat, niin jo hyvin pienilläkin parannuksilla on huomattavan suuret kansantaloudelliset vaikutukset. Avoimuus on olennaista julkisten hankintojen toimivuudelle. Tehokkaasti toimivat markkinat edellyttävät avointa ja helposti saatavilla olevaa tietoa tarjousten vertailussa käytetystä hinnasta tai muusta tekijästä. Tämän takia on suositeltavaa harkita Supervisorin kaltaisen internet-sivuston rakentamista myös Suomessa.

6 Johtopäätökset ja suositukset

6.1 Johtopäätökset

Datamassojen hyödyntäminen tulee läpäisemään koko yhteiskunnan aina liiketoiminnasta julkisiin palveluihin ja teollisuudesta yksilön käyttäytymiseen. Massadatan hyödyntämisen eturintamassa Suomessa ovat isot yritykset sekä toimialoista rahoitus- ja vakuutustoimintaa harjoittavat ja energia-alan yritykset ja informaatio- ja viestintäsektorilla toimivat yritykset. Massadatan potentiaalia nähdään erityisesti teollisuudessa, logistiikassa, julkisten palveluiden tuotannossa ja ympäristöasioissa. Yrityksissä massadatan liiketoimintapotentiaalin uskotaan materialisoituvan parempien päätösten, uusien ja laadukkaampien tuotteiden ja asiakaskokemuksien paranemisen kautta. Monilla sektoreilla massadatan hyödyntäminen on kuitenkin vielä varsin varhaisessa vaiheessa, eikä yrityksissä nähdä massadatan tarjoamia liiketoimintamahdollisuuksia. Toiseksi merkittävin rajoite massadatan hyödyntämiselle yrityksissä on tietotaidon puute.

Myös kaupungit ja kunnat ovat uusien haasteiden edessä massadatan kysynnän kasvaessa. Kaupungit näkevät massadatasta olevan eniten hyötyä palveluiden tuottamisessa ja päätöksenteossa. Käytännössä massadataa käytetään kuitenkin vielä niin vähän, että on vaikea ennustaa, millä julkisen sektorin aloilla sitä tullaan käyttämään eniten. Tutkimukset ovat nostaneet esiin sekä massadatan liiketoiminta- että tuotekehityspotentiaalin, mutta suurin osa kaupungeista ja yrityksistä ei ole tarttunut tähän potentiaaliin. Vaikka kaupungit näkevät roolikseen kaupungeissa syntyvän datan avaamisen, paikallisten innovaatioekosysteemien rakentamisen ja sitä kautta massadatan käytön mahdollistamisen, ei tietovarantoja ole vielä avattu ja hyödynnetä täysimääräisesti. Tietojärjestelmiä kaupungeissa ja kaupunkiseuduilla on sadoittain, ja tiedonhallinta on hyvin hajanaista. Datan avaaminen tulee vaatimaan kaupunkien eri tietopalveluiden ja dataa keräävien ja tietoa hyödyntävien tahojen tiivistä yhteistyötä.

Datan avoimuus herättää sekä kiinnostusta että huolta. Haastatteluissa ja työpajoissa nähtiin olennaisena kiinnittää huomiota datan omistajuus-, pääsy- ja tietosuojakysymyksiin. Kaupungin tulee yhtäältä huolehtia avaamansa datan tietosuojasta ja omistajuuskysymyksistä ja toisaalta tarjota alusta metadatan välittämiseksi.

Kansainväliset esimerkit osoittavat, että massadatan hyödyntäminen tietosuojalainsäädäntöä noudattaen on paitsi synnyttänyt uusia innovatiivisia ratkaisuja, myös joissain tapauksissa johtanut julkisen sektorin toimijoita uudelleen arvioimaan tehtäviään ja rooliaan. Esimerkiksi monimutkaista tietosuojalainsäädäntöä ymmärtämään auttavat työkalut yksinkertaistavat abstraktia lainsäädäntöä ja käsitteitä helpommin ymmärrettävään muotoon. Tällaiset työkalut voivat edistää muun muassa datan uudelleenkäyttöä sekä yksityisellä että julkisella sektorilla ja helpottaa ja automatisoida tietosuojaviranomaisten omaa työtä. Kansainväliset esimerkit massadatan hyödyntämisestä myös kyseenalaistavat tietosuojan absoluuttisena oikeutena. Jossain tilanteissa periaatepunninta tietosuojan ja muiden julkisten intressien välillä on tarpeen, kun otetaan huomioon yhteiskunnalle syntyvät merkittävät hyödyt.

6.2 Suositukset

Yritykset: Yritysten osalta massadatan hyödyntämisen suurimpina esteinä ovat se, ettei massadatan hyödyntämisestä syntyviä liiketoimintamahdollisuuksia täysin ymmärretä, sekä datan käsittelyyn ja analyysiin liittyvä tietotaidon puute. Ensiarvoisen tärkeitä toimenpiteitä ovat:

1. Resurssien suuntaaminen osaamisen riittävyyden varmistamiseksi datan käsittelyn, hallinnan ja analytiikan **koulutukseen**. Teknisten alojen lisäksi on syytä korostaa myös datatalouden yhteiskunnallis-taloudellisen koulutuksen merkitystä.
2. Resurssien suuntaaminen massadatan hyödyntämistä edistävään **tutkimus- ja innovaatiotoimintaan**. Teknisen tutkimuksen lisäksi tarvitaan datan hyödyntämiseen liittyvien laajempien yhteiskunnallis-taloudellisten mekanismien ymmärtämistä. Tämä on yksi edellytys onnistuneelle datapolitiikalle.
3. Datan hyödyntämiseen liittyvät **palvelukehityskokeilut** voivat toimia yrityksille oppimisalustoina massadatan liiketoimintapotentiaalin valjastamiseksi.

Kaupungit:

1. **Yhteiskehittelyä.** Kaupunkien tulee edelleen avata mahdollisuuksia yrityksille massadatan hyödyntämiseen eli tiedottaa, mitä tietoa on avoimesti hyödynnettävissä ja aktiivisesti edistää ja yhteiskehittää tiedon hyödyntämistä.
2. **Verkostoja teemoja ympärille.** Tiedonvaihtoa kaupunkien välillä ekosysteemien rakentamisessa kannattaa teemoittaa ja vuorovaikutusta syventää.
3. **Toimivat käytännöt esiin.** Julkisen sektorin toimijat voisivat tuoda selvemmin esiin massadatan käytön hyötyjä kertomalla hyvistä esimerkeistä ja käytännöistä. Kaupunkien ja kuntien esimerkit muun muassa taloudellisista säästöistä, lisääntyneestä asukkaiden osallistumisesta tai paremmasta päätöksenteosta rohkaisevat massadatan käyttöön. Ajankohtaisia aihioita massadatan hyödyntämisessä on erityisesti liikenteessä ja hyvinvointi- ja terveyssektorilla. Molemmat ovat kansallisesti merkittäviä strategisia painopistealueita kehittämisessä. Käytännön kokeiluja on jo liikkeellä, lisää kokeiluja ja kokemuksia tarvitaan.
4. **Osaamisen vahvistaminen.** Kaupunkien tulee varmistaa henkilöstön osaaminen datan käytössä ja avaamisessa. Erityisesti tausta- ja käytettävyydestietojen kuvaamisella on olennainen merkitys datan käytettävyyden kannalta. Tiedon hyödyntämisessä osaaminen on usein asiantuntijatasolla. Kaupunkien virkamies- ja poliittisen johdon ymmärrystä massadatan potentiaalista on laajennettava ja syvennettävä.

Julkinen sektori: Julkisella sektorilla olisi syytä huolellisesti **arvioida kansainvälisten innovatiivisten massadatan hyödyntämiseen liittyvien ratkaisujen soveltuvuutta ja kehittymahdollisuuksia Suomessa**. Uusien työkalujen kehittäminen luo mahdollisuuksia tehostaa julkisen sektorin toimintaa, parantaa sen palveluita ja lisätä päätöksenteon ja resurssien allokoinnin avoimuutta sekä mahdollisesti edistää myös massadatan hyödyntämistä (esim. datan uudelleen käyttöä eri tarkoituksiin) sekä yksityisellä että julkisella sektorilla.

Massadataekosysteemin kehittäminen: Omadata on kriittisessä asemassa massadataekosysteemin kehittämisessä. Henkilöistä ja heidän toiminnastaan kerättyä dataa voi sisältyä lähes mihin tahansa massadataan. Sen huono hallinta voi hidastaa verkkopohjaisten palvelujen kasvua ja massadatasta toivottujen hyötyjen saamista. Haasteiden ratkaisemiseen tarvitaan uusia toimintamalleja, kuten omadata-ekosysteemiä, missä painotus siirtyy lainsäädännöstä käyttäjien hallitsemiin datan hallinnan työkaluihin. Yksi lupaavimmista mahdollisuuksista omadata-työkalujen kehittämiseen ovat blockchain/lohkoketju-teknologiat, jotka sopivat hajautettuun autentikointiin, tiedon käsittelyyn, turvatarkoituksiin sekä arvon tai resurssien jakamiseen.

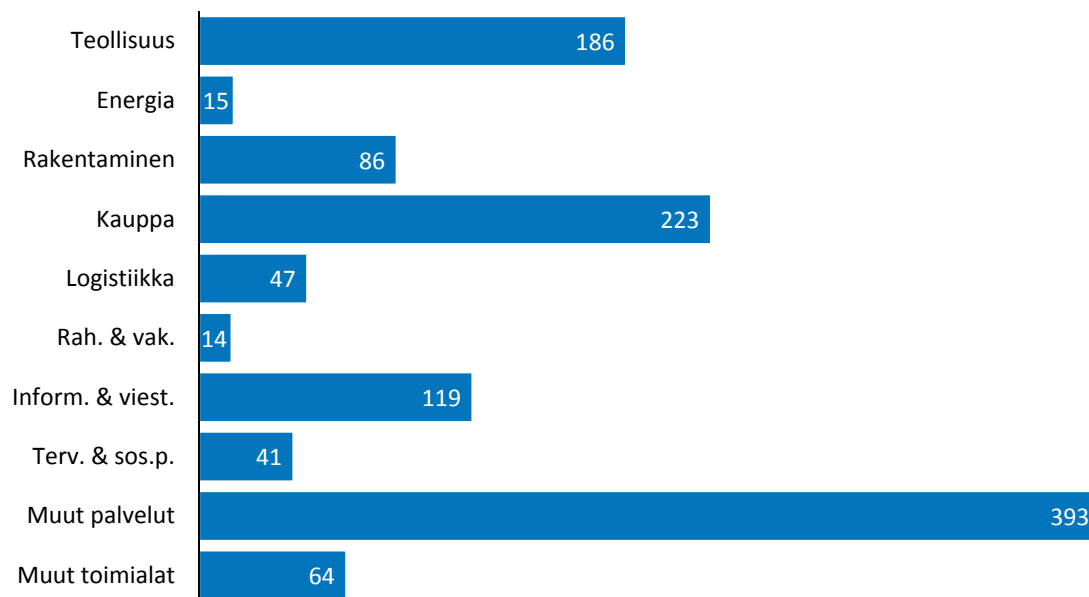
Lähteet

- Chen, Hsinchun – Chiang, Roger H. L. – Storey, Veda C. (2013). Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly.
- Data Protection Working Party (2013). Opinion 03/2013 on purpose limitation, 00569/13/EN. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf
- Davenport, Thomas D. (2014). Big data @ work. Harward Business Review Press.
- Furnas, Alexander (2013). Transparency Case Study: Public Procurement in the Slovak Republic. Saatavilla osoitteesta <http://sunlightfoundation.com/blog/2013/08/12/case-study-public-procurement-in-the-slovak-republic/>.
- Gutierrez, Daniel D. (2014). Inside BIGDATA guide to big data for finance. Dell & Intel.
- IDC (2015). European Data Market SMART 2013/0063. D6- First Interim Report. Oct 16th 2015.
- IDG Enterprise (2015). Big Data and Analytics Survey. <http://www.idgenterprise.com/report/2015-big-data-and-analytics-survey>
- LVM (2014). Big datan hyödyntäminen. Liikenne- ja viestintäministeriön julkaisuja 20/2014.
- LVM (2016). Massadataan kohdistuva sääntely. Digitaalisen liiketoiminnan kasvuympäristökärkihanke. LVM:n taustaraportti (julkaisematon).
- Mattila, Juri – Seppälä, Timo (2015). Laitteet pilveen – vai pilvi laitteisiin? Keskustelunavauksia teollisuuden ja yhteiskunnan digialustojen uusista kehitystrendeistä. ETLA Raportit No 44. <http://pub.etla.fi/ETLA-Raportit-Reports-44.pdf>
- Mayer-Schönberger, Viktor – Cukier, Kenneth (2013). Big Data – A Revolution That Will Transform How We Live, Work and Think. John Murray.
- OECD (2015). Data-Driven Innovation: Big Data for Growth and Well-Being. OECD Publishing. Paris.
- Peuralainen, Johanna – Muttilainen, Vesa (2015). Korruption riskikohteet 2010-luvun Suomessa. Poliisi-ammattikorkeakoulun raportteja 115.
- Poikola, Antti – Kuikkaniemi, Kai – Kuittinen, Ossi (2014). Open Knowledge Finland: MyData – johdatus ihmiskeskeiseen tiedon hyödyntämiseen. <http://www.lvm.fi/-/my-data-johdatus-ihmiskeskeiseen-henkilötiedon-hyodyntamiseen-842175>
- Purdy, Mark – Davarzani, Ladan (2015). The Growth Game-Changer: How the Industrial Internet of Things can drive progress and prosperity. Accenture.
- TNS Political & Social – Flash Eurobarometer 374 (2014). Business' Attitudes Towards Corruption in the EU. February 2014.
- Turck, Matt – Hao, Jim (2016). The 2016 Big Data Landscape. FirstMark Capital.
- Wolkowitz, Eva – Parker, Sarah (2015). Big Data, Big Potential: Harnessing Data Technology for the Underserved Market. Center for Financial Services Innovation.

Liite 1

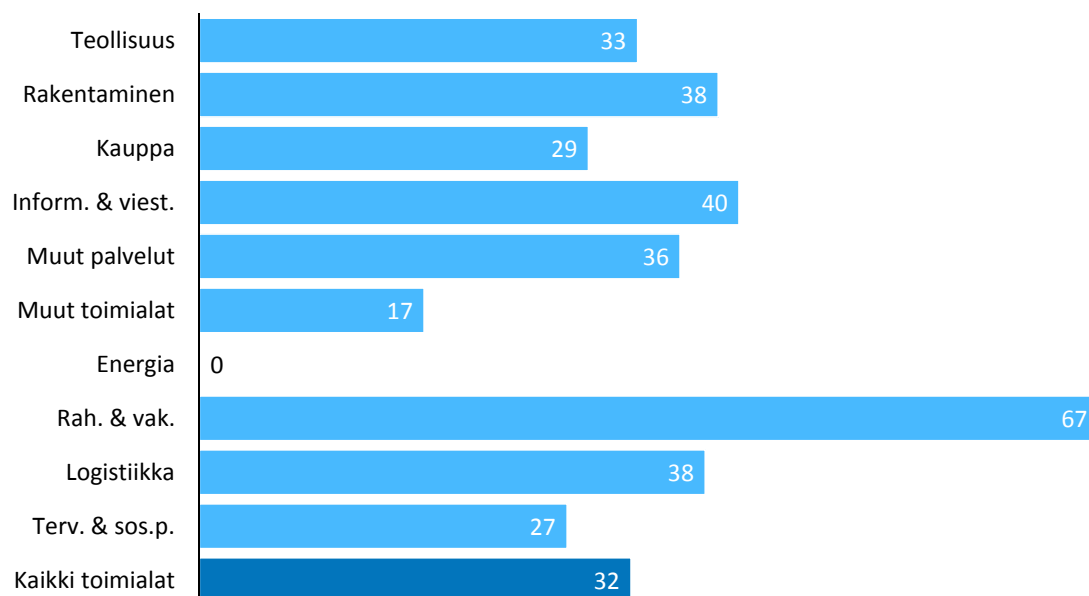
Yrityskysely: kuvioita

Kyselyyn vastanneiden lukumäärä toimialoittain



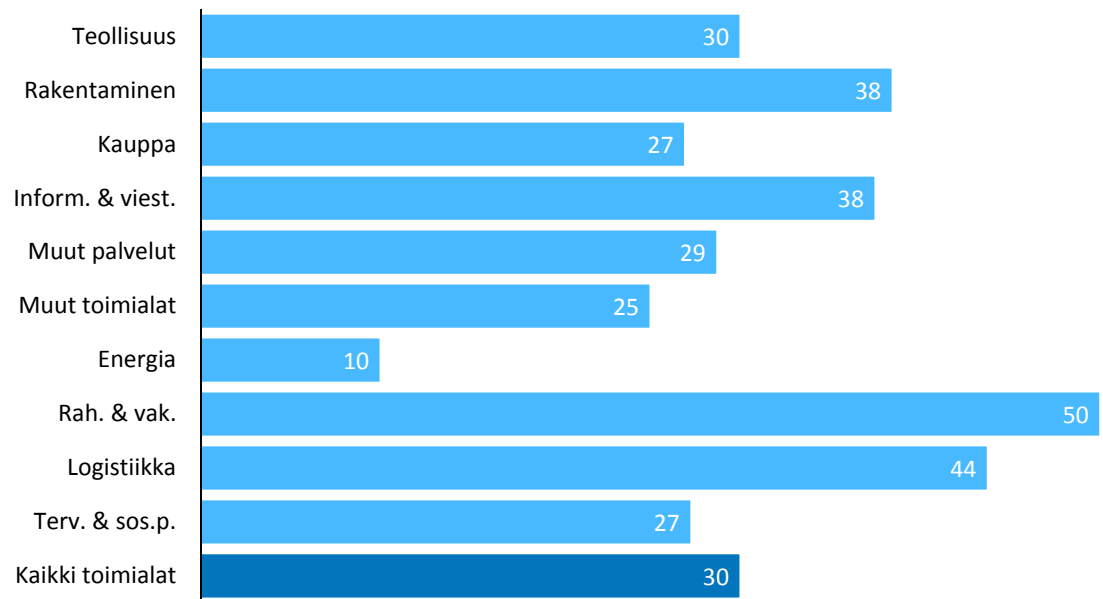
Lähde: ETLAn kysely, marraskuu 2015.

Massadatan käyttö vaikuttaa nyt/lähtulevaisuudessa melko tai hyvin paljon:
Liikevaihdon kasvu, % yrityksistä



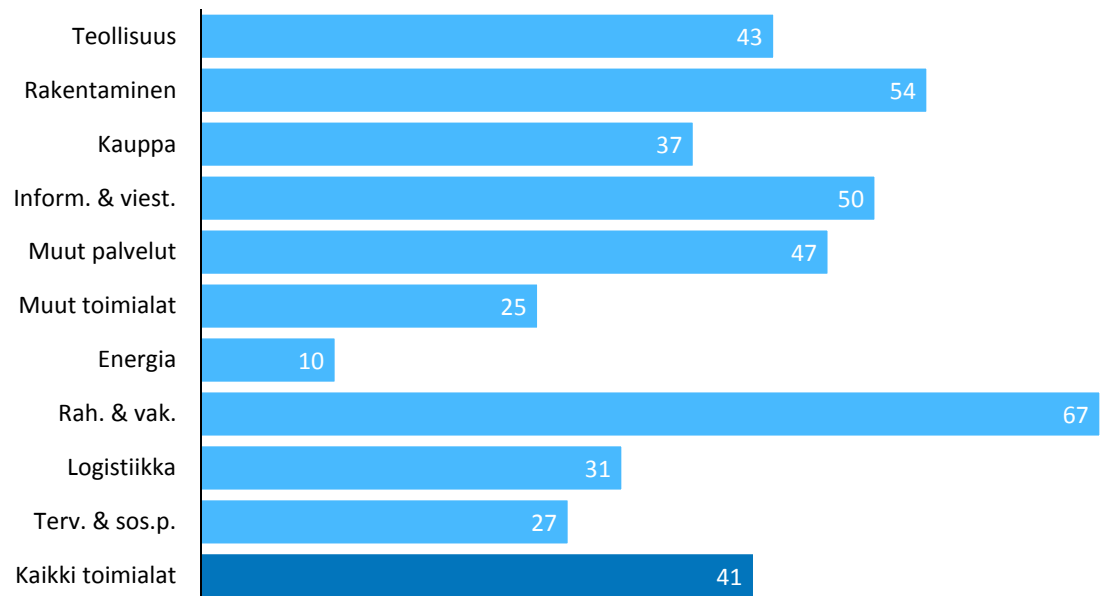
Lähde: ETLAn kysely, marraskuu 2015.

Massadatan käyttö vaikuttaa nyt/lähitulevaisuudessa melko tai hyvin paljon:
Kustannusten säästö, % yrityksistä



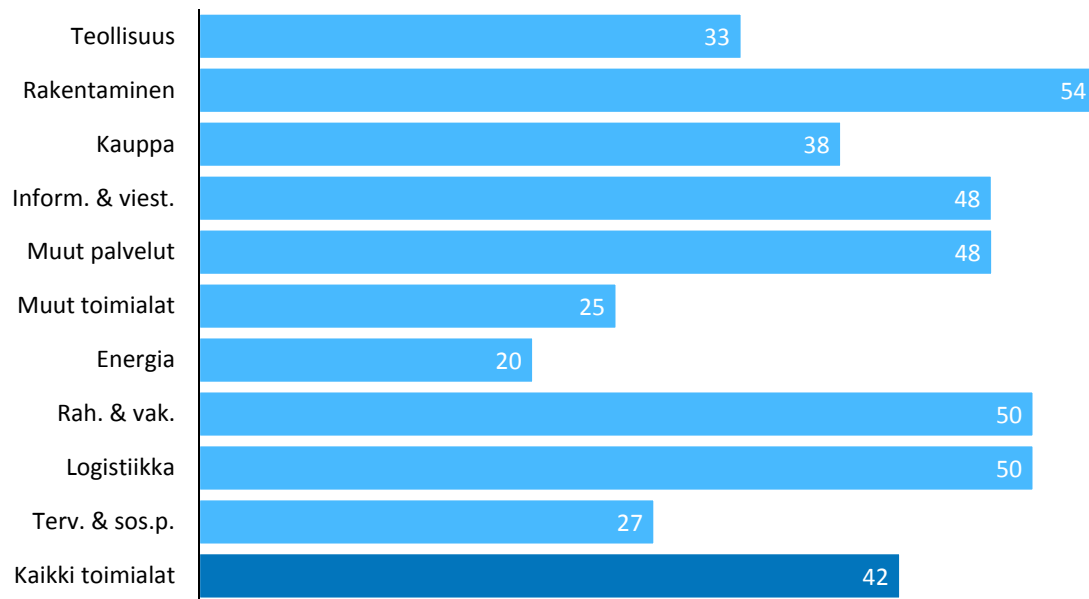
Lähde: ETLAn kysely, marraskuu 2015.

Massadatan käyttö vaikuttaa nyt/lähitulevaisuudessa melko tai hyvin paljon:
Tuotteiden laadun paraneminen, % yrityksistä



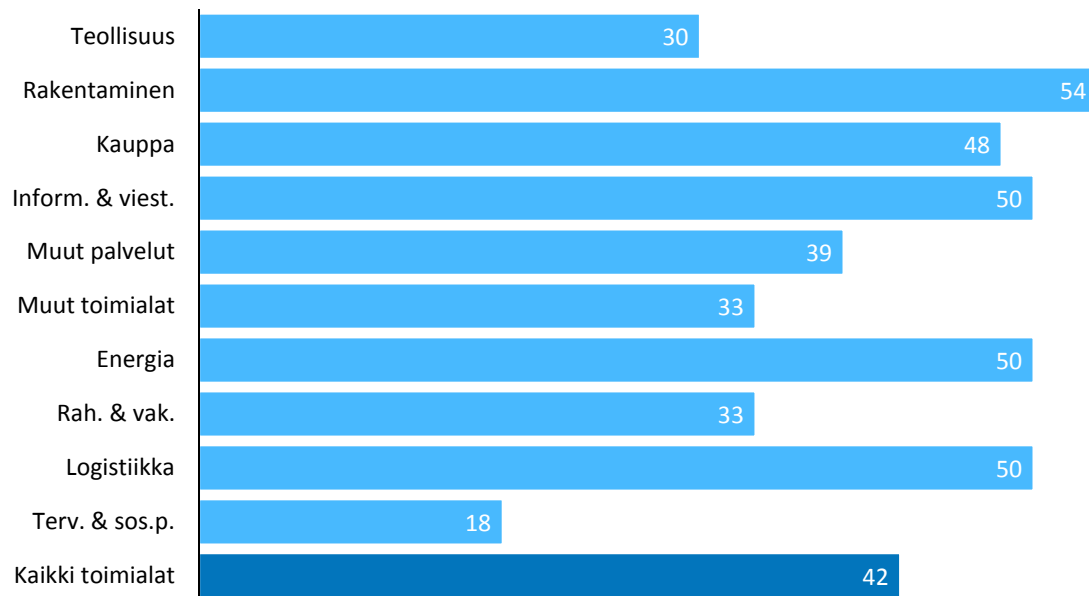
Lähde: ETLAn kysely, marraskuu 2015.

Massadatan käyttö vaikuttaa nyt/lähtulevaisuudessa melko tai hyvin paljon:
Kilpailuaseman paraneminen, % yrityksistä



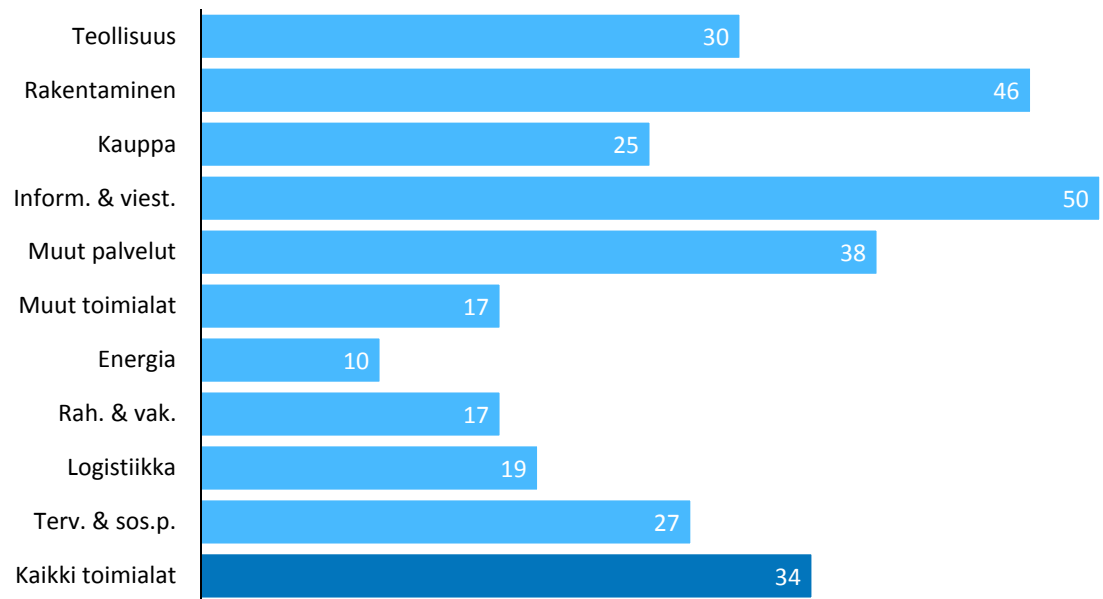
Lähde: ETLAn kysely, marraskuu 2015.

Massadatan käyttö vaikuttaa nyt/lähtulevaisuudessa melko tai hyvin paljon:
Asiakaskokemuksen paraneminen, % yrityksistä



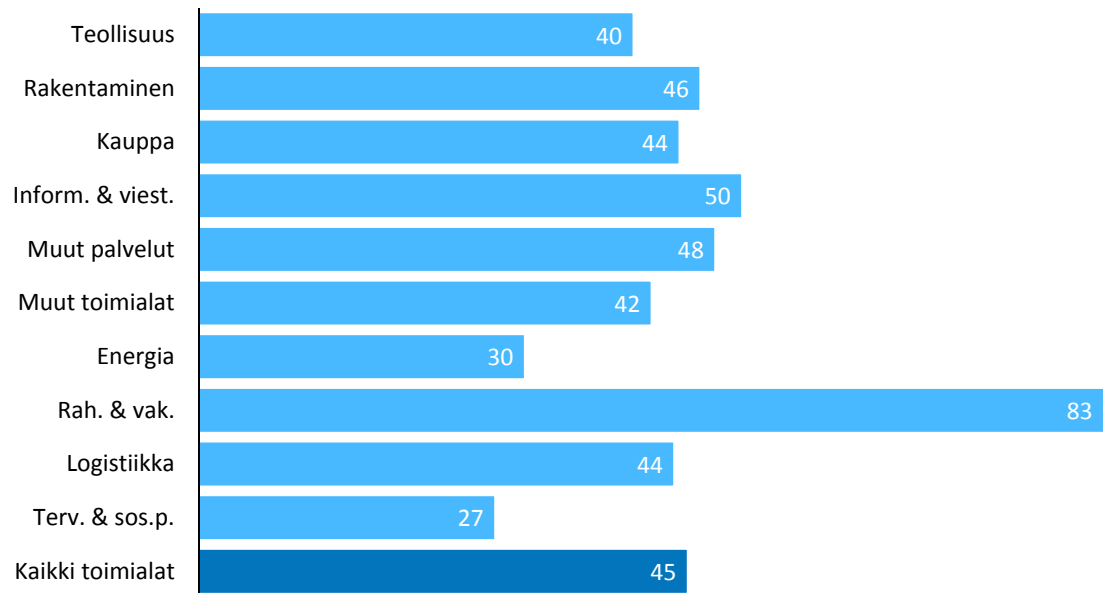
Lähde: ETLAn kysely, marraskuu 2015.

Massadatan käyttö vaikuttaa nyt/lähitulevaisuudessa melko tai hyvin paljon:
Uusien tuotteiden syntyminen, % yrityksistä



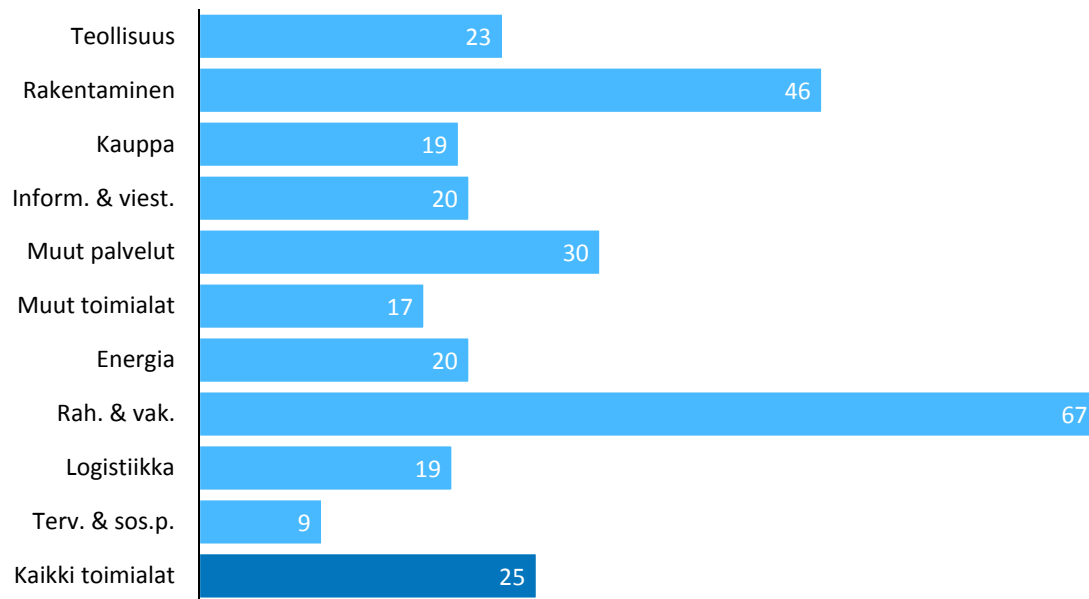
Lähde: ETLAn kysely, marraskuu 2015.

Massadatan käyttö vaikuttaa nyt/lähitulevaisuudessa melko tai hyvin paljon:
Paremmat, tietoon perustuvat päätökset, % yrityksistä



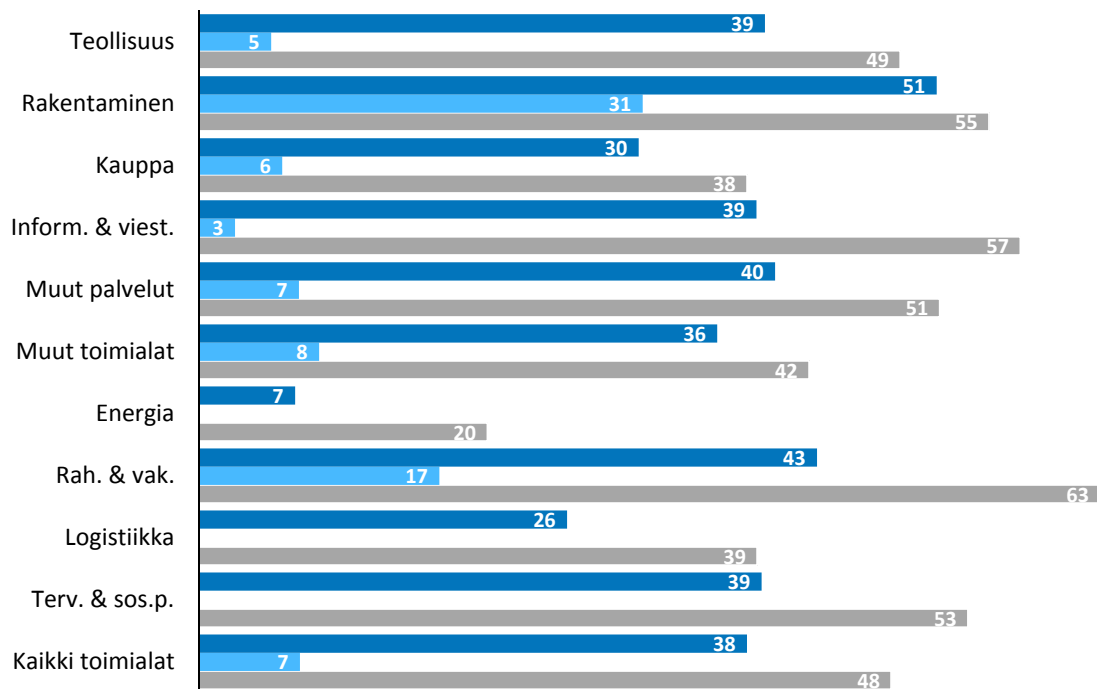
Lähde: ETLAn kysely, marraskuu 2015.

Massadatan käyttö vaikuttaa nyt/lähitulevaisuudessa melko tai hyvin paljon:
Liiketoimintariskien vähentyminen, % yrityksistä



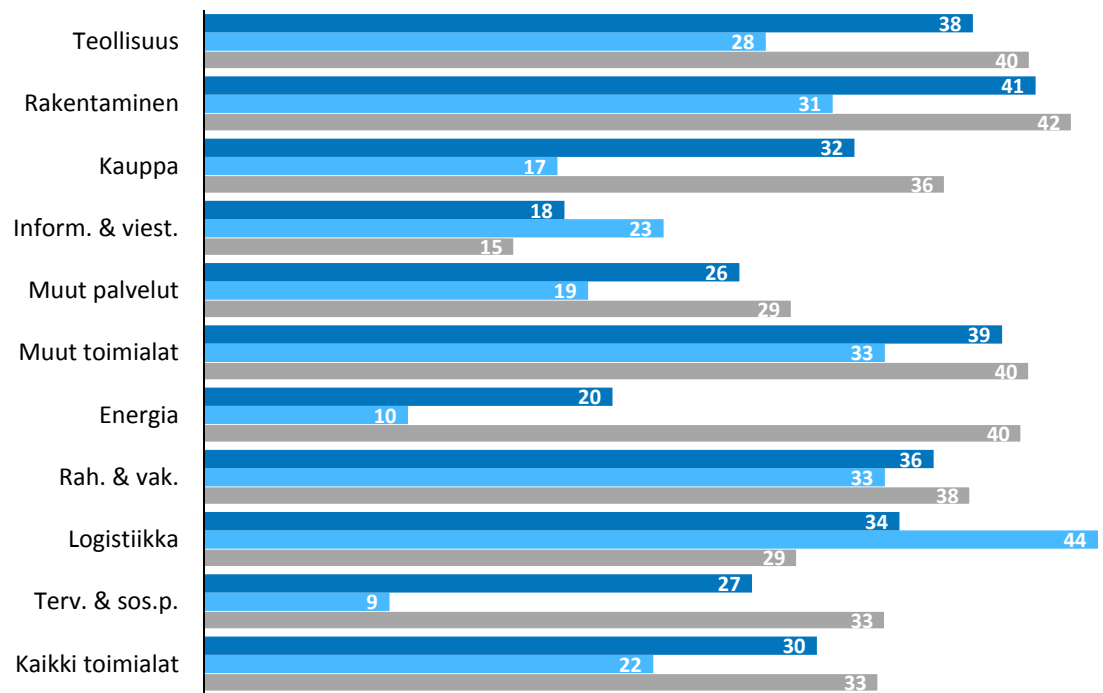
Lähde: ETLAn kysely, marraskuu 2015.

Rajoittaa tai estää massadatan hyödyntämistä liiketoiminnassa melko tai hyvin paljon:
Ei ole tarvetta käyttää massadataa liiketoiminnassa, % yrityksistä



Lähde: ETLAn kysely, marraskuu 2015.

Rajoittaa tai estää massadatan hyödyntämistä liiketoiminnassa melko tai hyvin paljon:
Datan käsittelyyn ja analyysiin liittyvän tietotaidon puute, % yrityksistä



Lähde: ETLAn kysely, marraskuu 2015.

Liite 2

Lista esimerkeistä koskien innovatiivisia ratkaisuja massadatan hyödyntämisessä

Alla oleva taulukko sisältää tiivistelmän potentiaalisista tapauksista, jotka voisivat toimia esimerkkeinä innovatiivisista sovelluksista henkilötietojen massadatan ja/tai avoimen datan alueella. Lista on koottu lähettämällä pyyntö noin 30 eurooppalaiselle asiantuntijalle. Esimerkkien alustava valinta esiteltiin hankkeen ohjausryhmälle 09.12.2015.

Esimerkin nimi	Alkuperämaa	Tiivistelmä ja potentiaalinen arvo	esimerkin yhteydet toimijoiden välillä (H = julkinen hallinto, L = liike-elämä K = kansalainen)					
			H2H	H2L	H2K	L2L	L2K	K2K
<u>Supervizor</u>	Slovenia	Supervizor on verkkosovellus, joka tarjoaa käyttäjille informaatiota julkisen sektorin toimielimien liiketoimista. Julkisen sektorin toimielimiin, jotka ovat suoria tai epäsuoria budjetin käyttäjiä, kuuluvat mm. lainsäädäntöelimet, oikeuslaitos, autonomiset ja itsenäiset valtion toimielimet, kunnat ja kuntayhtiöt, julkiset instituutit, julkinen rahoitus ja julkiset virastot. Sovellus on suunniteltu ja toteutettu Slovenian valtion korruption vastaisen komission ja yksityisen asiantuntijan yhteistyönä. Relevantit tiedot on saatu toimielimiltä, jotka ovat auttaneet myös informaation tulkinassa. Apuna on ollut Slovenian valtionvarainministeriö, Slovenian valtion julkisten maksujen hallinto UJP, Slovenian valtion julkisen oikeudellisen tiedon virasto AJPES. Henkilötietojen tietosuojaherkkyden vuoksi osa tiedoista on pitänyt poistaa hakukelpoisten tietojen joukosta henkilötietosuojasta vastaavan viranomaisen määräyksestä		x	x			
<u>AEGLE-projekti</u>	EU	Aeglen tavoitteena on aikaansaada eurooppalainen liiketoiminnan ekosysteemi terveydenhuollon sidosryhmille, toimialalle ja tutkijoille, jonka sisällä luodaan uudenlaista tietämystä. Tarkoituksena on tarjota pilvi- ja HPC-datapalveluita ja tukea terveyttä edistäviä uusia tuotteita.				x		

Esimerkin nimi	Alkuperämaa	Tiivistelmä ja potentiaalinen arvo	esimerkin yhteydet toimijoiden välillä (H = julkinen hallinto, L = liike-elämä K = kansalainen)					
			H2H	H2L	H2K	L2L	L2K	K2K
<u>Italialaiset koulut/vakit uinen virka</u>	Italia	Esimerkki uudelleenmuokatusta henkilötieto-datasta, jonka henkilötietosuojasta vastaava viranomais Italiassa on sallinut. Tietohauulla voidaan järjestää informaatio sen mukaan missä alueella tai provinssissa henkilöllä on suurempi mahdollisuus saada vakituinen virka perustuen mm. henkilön sijoittumiseen tietystä joukossa ja muiden hakemusten lukumäärään. Huomioitavaa on se, että palvelu käyttää henkilötietoja ja on tästä huolimatta viranomaisen nimenomaisesti sallima. Perusteena on se, että palvelun tarkoitus on yhtenevä sen tarkoituksen kanssa, joka alkuperäisellä tietokokoelmalla on ollut.			x			
Henkilötieto ja sisältävän datan uudelleenkäytön työkalu	Hollanti	Työkalun avulla julkisen sektorin toimielimet voivat arvioida, voidaanko henkilötietoja, jotka on kerätty tarkoitusta A varten, käyttää myös tarkoitukseen B. Toimenpiteellä on suuri arvo erityisesti Alankomaissa, jossa eri julkisen sektorin toimijat haluaisivat jakaa tallentamaansa tietoa (usein sijaintitietoihin perustuvaa tietoa) myös ulkopuolisten organisaatioiden kanssa, mutta tietosuojasysteistä asiassa ei uskalleta edetä.	x	x	x			
Almeren kunta (Straatkubus)	Hollanti	Henkilötietojen linkittäminen ja prosessointi sekä jakaminen sidosryhmille (valtion organisaation sisällä ja ulkopuolelle).	x	x	x			
Hollannin tilastoviranomainen (CBS)	Hollanti	Alkuvaiheessa oleva projekti, jossa tilastoista vastaava virasto toimisi valtuutettuna osapuolena henkilötietojen prosessorijana. Muiden viranomaisten keräämästä henkilötieto-datasta muokataan versio, joka noudattaa henkilötietosuojasäännöksiä.	x	x	x			
<u>My Care Pathways</u>	Ruotsi	Potilaat voivat sallia kolmannen osapuolen pääsyn dataan.	x	x	x	x	x	

Esimerkin nimi	Alkuperämaa	Tiivistelmä ja potentiaalinen arvo	esimerkin yhteydet toimijoiden välillä (H = julkinen hallinto, L = liike-elämä K = kansalainen)					
			H2H	H2L	H2K	L2L	L2K	K2K
Nugg Ad Smart Audience Platform	Saksa	Älykästä mainonnan kohdistamista. Monikanavaista yleisöön kohdistamista perustuen erilaisiin datoihin, Smart Data-hallintaan, sekä yksilöllisesti suunniteltuihin julkaisijan ja mainostajan kohderyhmiin. Nugg Ad:in etuna on pitkäaikainen kohderyhmä-asiantuntemus ja innovatiivinen teknologia, joka perustuu koneiden reaaliaikaiseen oppimiseen. Yhteistyökumppanit hyötyvät tehokkaasta digitaalisesta mainonnasta ja laajasta kohderyhmä-asiantuntemuksesta. Vuonna 2014 Nugg Ad:in tuotteille myönnettiin ePrivacyseal ePrivacyconsult GmbH:n toimesta. Nugg Ad anonymisoi kohderyhmiensä tiedot ja näin ollen palvelun käyttäjien ei tarvitse huolehtia henkilötietolainsäädännöstä. Varmistaakseen tiedon maksimaalisen suojaamisen Nugg Ad noudattaa eurooppalaisia itsesääntelystandardeja ja ns. EDAA Trust Seal -sertifikaattia. Nugg Ad, joka on perustettu 2006, on alansa pioneeri ja johtava tarjoaja ennustavan kohdistamisen saralla. Nugg Ad mahdollistaa yhteistyökumppaniensa yksilöidysti kohdistetun, ryhmäsidonnaisen digitaalisen mainonnan näytöille, videoille ja mobiilimainontaan – suoran myynnin tai ohjelmoidun alustan.				x	x	
Clarus secure cloud computing services	EU	Viitekehys käyttäjäkeskeiselle yksityisyydensuojalle ja turvallisuudelle pilvessä.				x	x	x

ISSN PDF 2342-6799
ISBN PDF 978-952-287-250-0
Snellmaninkatu 1, Helsinki
PL 23, 00023 VALTIONEUVOSTO
p. 0295 16001
f. 09 1602 2165

